WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

EMPIRICAL APPROACHES TO THE PROBLEM OF AGGREGATION OVER INDIVIDUALS

by
Thomas M. Stoker

Massachusetts Institute of Technology

WP#3549-93-EFA March 1993

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

EMPIRICAL APPROACHES TO THE PROBLEM OF AGGREGATION OVER INDIVIDUALS

by
Thomas M. Stoker

Massachusetts Institute of Technology

WP#3549-93-EFA March 1993

EMPIRICAL APPROACHES TO THE PROBLEM OF AGGREGATION OVER INDIVIDUALS

By Thomas M. Stoker

June 1992, Revised March 1993

Forthcoming in *Journal of Economic Literature*

* Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139. This work was funded by National Science Foundation Grant SES 9122886. I have received numerous valuable comments from many individuals, comprising a list that is too long to acknowledge here. However, special thanks go to John Heaton and Dale Jorgenson, who provided comments that materially improved this work through each version. I also wanted to acknowledge a fantastic job by John Pencavel and every one of the reviewers, whose comments also materially improved this article.

EMPIRICAL APPROACHES TO THE PROBLEM OF AGGREGATION OVER INDIVIDUALS

1. Introduction

One of the most challenging features of tracking economic activity over time is assessing the impact of the changing composition of the economic players. In the United States, the decline in the typical size of households, the baby boom - baby bust cycles, the changing age structure of the population and the migration of households to southern climates provide examples of such changes. The shift of production from manufacturing and agriculture to service industries, and the continuing infusion of high technology throughout many areas provide examples of how the nature of production has varied. In most if not all aspects, the U.S. economy of the 1990's is considerably different from the U.S. economy of the 1950's, the 1960's and the 1970's.

If the economic reactions of such different kinds of players were nonetheless quite similar, then compositional effects on aggregate economic activity would be minor. In this case, compositional changes over time would amount to a relabeling of the economic players that is not associated with any real behavioral differences. However, for the U.S. or any other actual economy, this possibility is in conflict with casual observation and virtually all studies of disaggregated, microeconomic data.

Consider the needs for food and clothing of a large family relative to a small family at the same budget level, or of a poor family relative to a wealthy one. Consider the needs for health care of a young couple compared with an elderly couple, or more generally, the needs for current saving or having previously accumulated wealth. Consider the different concerns of a capital-intensive manufacturing company relative to a labor-intensive service provider, in trying to make plans for expansion or other new business investment. In broader terms, to the author's knowledge there are no studies of disaggregated, micro level data that fail to find strong systematic evidence of individual differences in economic behavior, whether one is concerned with demographic differences of families or industry effects in production. Entire empirical methodologies have been developed to account for systematic individual differences in micro level surveys, such as the modeling of fixed or random effects in panel data.

The presence of these kinds of differences have one strong implication for aggregate economic activity. Namely, it matters how many households are large or small, how many are elderly and young, and how many companies are capital-intensive or labor-intensive. Such heterogeneity of concerns and reactions are an essential feature of the overall welfare impacts of changes in food prices, the overall impacts of interest rates on savings, or the impact of an investment tax credit. It is difficult to conceive of an important question of economic policy that does not have a distributional component, or a differential impact on economic players. It is likewise hard to envision how the impacts of relative price changes or of real income growth could be adequately summarized over time without some attention to the composition of the economy.

Concerns over the issues raised by compositional heterogeneity of data on groups, such as data on economy-wide aggregates over time, are summarized under the heading of the "problem of aggregation over individuals." Over the past decade, various approaches have been developed to account for compositional heterogeneity in empirical modeling of aggregate data, and the purpose of this survey is to discuss this work.

To spell out the context of our survey in more detail, it is useful to differentiate the three major approaches to empirical modeling of aggregate data; 1) modeling aggregate data alone, including the representative agent approach, 2) modeling individual economic behavior alone, or microsimulation, and 3) joint modeling of individual and aggregate level data. These approaches differ in terms of their treatment of the problem of aggregation over individuals, and we now discuss them in turn.

The first approach is the econometric modeling of aggregate data series alone, where one asserts the existence of a stable model among aggregates, and then fits the model statistically. This approach is motivated as a first-cut method of studying aggregate data patterns, for the purpose of forecasting or getting rough insights on the interplay of macroeconomic variables for the analysis of economic policy. One version of this approach includes traditional macroeconomic equations, which are specified and estimated in an ad hoc fashion, with regressor variables included to represent the major economic influences on the macroeconomic variable under study. Examples include a standard Keynesian consumption function, an ad hoc money demand equation, or an accelerator-style investment equation, all familiar from macroeconomic textbooks. One can likewise include in this category the growing literature on pure time series analysis of macroeconomic data, where economy-wide aggregates are analyzed as the result of fairly stable stochastic processes (with unit roots and other concepts of cointegration used as the primary focus).

While we will survey some work that studies compositional influences in an ad hoc fashion, these purely statistical approaches are not well grounded in any model of (individual) economic behavior, and amount to making involved inferences solely from correlations between aggregate data series. Use of economic effects estimated in this way for policy analysis, or use of such equations for prediction, amounts to extrapolation of the recent past data patterns into the future, with no foundation relative to the behavior of the individual economic actors. This kind of traditional macroeconomic modeling amounts to a purely statistical approach to aggregate data series, motivated as the simplest method of parsimoniously summarizing interactions among aggregate variables.

This category also includes the tightly parameterized econometric models of individual consumer and firm decision making under uncertainty, that are related to aggregate data under the guise of a "representative agent." These models, the workhorses of modern real business cycle theory, treat economic aggregates as though they necessarily obey the constraints of rational choices by a single decision maker, namely a "representative consumer" or "producer." This kind of modeling has proved a tremendous engine for the development of rational choice models over the last two decades, and their empirical application has developed into an ideology for judging aggregate data models. In particular, models are judged as "good" on the basis of whether they coincide with a sophisticated decision process for a single individual. Given the assumed "existence" of a representative agent, compositional issues are ignored by fiat.

There are various well-known settings in which the structure appropriate for a representative agent exists in aggregate data. For instance, in terms of demands for different goods, an aggregate preference relation exists if individuals have identical and homothetic preferences (Terence Gorman (1953)), or if the government is continuously redistributing income in an optimal way (Paul Samuelson (1956)). These kinds of conditions may seem far-fetched for any real-world economy, but they are representative of all "justifications" of the representative agent approach. In particular, no realistic conditions are known which provide a conceptual foundation for ignoring compositional heterogeneity in aggregate data, let alone a foundation for the practice of forcing aggregate data patterns to fit the restrictions of an individual optimization problem.

There is a vast and growing literature on what is wrong with the representative agent approach, that is well surveyed in Alan Kirman's (1992) stinging criticism. While we do note some implications of ignoring heterogeneity in aggregate data as part of our motivation, our current purpose is to discuss methods of incorporating individual differences in aggregate data models. As such, a broad posture of our exposition is that recent developments underscore how a "representative agent" is not necessary for incorporating economic restrictions in aggregate data. Taken at face value, representative agent models have the same value as traditional, ad hoc macroeconomic equations; namely they provide only statistical descriptions of aggregate data patterns, albeit descriptions that are straight jacketed by the capricious enforcement of restrictions of optimizing behavior by a single individual. Without attention to aggregation, one can only be skeptical about using empirical results from a model that is motivated solely by the phrase "assume a representative agent."

A natural reaction to the difficulties of ignoring heterogeneity is to carry out all behavioral modeling at the level of individual agents. The second approach to modeling aggregate data is microsimulation, which takes this posture to its extreme. In particular, this approach begins with a full model of the behavior of each different type of individual in the population, estimated with survey or panel data on individuals. Aggregate values are then simulated by adding up across all individuals. Examples of microsimulation models include the Joint Committee on Taxation's (1992) model for simulating tax policy impacts, and various models of appliance choice and energy demand, such as those described by Thomas Cowing and Daniel McFadden (1984).

Microsimulation models have the potential for the most realistic representation of aggregate data movements - an adequate model of the behavior of each kind of economic player would represent the full behavioral foundation underlying economic aggregates. The drawback to this kind of model is not in its foundation, but rather in practical implementation. Supposing that a complete individual model can be characterized without difficulty (and this is a huge supposition), microsimulation involves carrying out a separate simulation of each individual's behavior. Consequently, exogenous and/or predetermined variables need to set for each individual, as well as starting conditions when the individual models are dynamic. With a substantive accounting for individual differences, there is virtually unlimited flexibility in the application of microsimulation models, but the simulation process becomes virtually intractable to carry out.

Because the results for aggregated variables are dependent upon precisely how the individual simulations are specified, the sheer scale of possible inputs precludes any meaningful understanding of the primary influences on aggregate data movements. In particular, with the exception of the work of James Heckman and James Walker discussed below (Section 4.6), there have been no conclusive comparisons of aggregate tracking performance between microsimulation models and statistical models of aggregate data alone. For microsimulation models even the simplest form of aggregate validation is either difficult or impossible. This sobering feature of the microsimulation approach is clearly evidenced in the careful analysis of microsimulation models of energy demand of Cowing and McFadden (1984).

The third approach to modeling aggregate data, the subject of our survey, is to adopt a framework that permits individual data and aggregate data to be modeled under one consistent format. In particular, an individual model is specified together with assumptions that permit an aggregate model to be formulated that is consistent with the individual model. This approach models the comparability of individual behavioral patterns and aggregate data patterns, removing any mystery induced by the one-sided focus of studying aggregate data alone or individual data alone as in the other approaches.

The overall aim for models that account for aggregation over individuals is to account for individual heterogeneity as in the microsimulation approach, as well as give a tractable, parsimonious model for aggregate data. This compromise between the other approaches is typically achieved by using individual level equations that are restricted to accommodate aggregation, together with information on the distributional composition of the population. These added restrictions can be tested with individual and (sometimes) aggregate data, and such testing is necessary for a full validation of this kind of model.

There are clear advantages to such micro-macro models, which are useful to list at the outset. First, any restrictions on behavior applicable at the individual level model can be applied in a consistent fashion to the aggregate model. The parameters of individual level equations appear in the aggregate level model, and restrictions on those parameters (from individual optimizing behavior) are applicable at both levels. Second, simultaneous modeling of both individual and aggregate level data permits pooling of both kinds of data, which broadly allows heterogeneity to be characterized by observed individual differences in behavior. Finally, the results of estimating such a model are applicable to a wide range of applied questions; the individual level model can be used to measure distributional effects, and the aggregate level model used to simulate or forecast future aggregate data patterns. By construction, simulations of individual level equations are consistent with simulations of the aggregate level equations.

We have introduced these issues in a somewhat abstract fashion to set the stage. The elucidation of the principles involved in building models that account for aggregation, as well as recent examples of these kinds of models, comprise the subject of our survey. For a bit of a road map, we begin by a simple discussion of the issues raised by individual heterogeneity for equations fit to aggregate data. We then discuss some theoretical ideas that clarify how individual level models can differ from aggregate data patterns, as well as spell out what constitutes a well grounded, interpretable aggregate level model. This sets the stage for our survey of applied work, which is somewhat of a collage of different aspects of modeling aggregation over individuals. We begin with statistical methods of assessing distributional effects in aggregate data, that, while crude, point up interesting interactions between individual heterogeneity and aggregate dynamics. We then cover recent work in demand analysis, where micro-macro modeling has been developed most fully. Following this are sections discussing aggregate equations and statistical fit, various aspects of dynamic modeling, models of market participation and recent work in microsimulation that is focused on aggregation issues. We survey a variety of problem areas to give broad coverage to work that connects individual and aggregate models, which has been used in empirical work or is closely relevant to empirical methods.

Our survey will deal with only a fraction of recent literature that addresses questions under the heading of "aggregation," and so it is necessary to mention some areas that are not covered. As mentioned above, we will not cover the myriad of arguments against representative agent modeling, nor the numerous theoretical results on what micro level assumptions can yield partial structure among aggregate variables. Kirman (1992) provides reasonable coverage of this literature. We are concerned with aggregation over individuals, and will not cover the construction of aggregates within individual decision processes, such as in the literature on commodity aggregates and two-stage budgeting, or in the literature on whether an aggregate "capital" construct is consistent a heterogeneous population of firms; a good starting point for these literatures is Charles Blackorby, Daniel Primont and Robert Russell (1978).

Moreover, we focus on macroeconomic variables that are averages or totals across an economy comprised of a large number of individual agents, and we presume that the definition of "individual agent" is sufficiently unambiguous to make sense out of applying an individual level model. For some of our discussion, we are interested in the recoverability of empirical patterns of individual level data from aggregate data series, and then the definition of "individual" is given from the context.[1] But for the applications of restrictions from rational behavior, such behavior is taken as appropriate for the "individual" so defined. For example, in the context of demand analysis, the "individual" is typically a household, and the application of integrability restrictions assumes that households are acting as a single rational planning unit. We do not cover the literature on whether decisions of multi-person households are made jointly, or are the aggregates of separate decisions made by the individual household members under a bargaining process. While the questions addressed in this literature overlap with some of our concerns, the setting of a two-to-seven member household is sufficiently different from a national economy to raise quite different issues. For example, ongoing income redistribution may be entirely feasible within the context of a single family, in a way that one would never consider applicable to a real-world economy. Good starting points for this literature include Robert Pollak (1985) and Pierre-André Chiappori (1988), among many others.[2]

2. Basic Issues of Heterogeneity and Aggregate Data

Traditional methods of modeling aggregation over individuals involve fairly strong linearity restrictions on the impacts of individual heterogeneity, and the theory we describe later indicates the role of such restrictions. For motivation of the basic problems, we first develop a feel for the issues raised by individual heterogeneity through some elementary examples.

Much of the work on aggregation has been developed in the context of analyzing commodity demands, and so we consider a simple static demand paradigm here. Suppose that our interest is in studying the demand for a commodity, as a function of prices and total expenditure budget, or "income." Suppose further that the vector of prices faced by all individual households are the same at time t, but that incomes vary across households and over time. We employ the following notation:

N_t: Number of households at time t; indexed by $i = 1,\ldots,N_t$.

p_t: Price (Vector) at time t.

y_{it}: Demand for the Commodity by household i at time t.

M_{it}: Total Expenditure Budget, or "Income" of household i at time t.

$y_{it} = f_i(p_t, M_{it})$: Demand Function of household i at time t.

$$\bar{y}_t = \frac{1}{N_t} \sum_i y_{it}: \text{ Average Demand}$$

$$\bar{M}_t = \frac{1}{N_t} \sum_i M_{it}: \text{ Average Income}$$

We are interested in how aggregate demand $\bar{y}_t$ relates to aggregate income $\bar{M}_t$ and price p_t.

A "representative agent" approach to studying aggregate demand could begin with a formulation of a "per capita" demand equation $\bar{y}_t = G(p_t, \bar{M}_t)$, presuming that mean demand $\bar{y}_t$ is determined solely by prices and mean income $\bar{M}_t$. This equation would be fit with aggregate data over time, by least squares or some other technique. The issues we discuss below are not affected by what estimation method is used, but arise solely because of the use of the aggregate income $\bar{M}_t$ alone to explain average demand $\bar{y}_t$. Consequently, for our examples we suppose that aggregate estimation reveals the true pattern between $\bar{y}_t$ and p_t, $\bar{M}_t$, without making reference to any particular estimation method.

Consider first a straightforward setting. Suppose that all households have identical homothetic preferences, so that given prices, all households allocate the same fraction of their income to the commodity. Demand for household i at time t is then expressible as

$$(2.1) \qquad y_{it} = b(p_t)\, M_{it} \,.$$

Here an additional dollar of income increases demand by $b(p_t)$ for any household.

In this case, aggregate demand is given by

$$(2.2) \qquad \bar{y}_t = b(p_t)\, \bar{M}_t \,.$$

This is a well defined, stable, interpretable relationship, which would be estimated with data on $\bar{y}_t$, p_t and $\bar{M}_t$ over time t. The reason for the stability of the relation is clear. Suppose that incomes change, inducing a change of $\Delta\bar{M}$ in $\bar{M}_t$. Each household adjusts their demand according to their common marginal effect $b(p_t)$, so aggregate demand changes by the marginal amount $b(p_t)\Delta\bar{M}$. The relation is interpretable because the aggregate marginal effect $b(p_t)$ is the marginal behavioral response $b(p_t)$ of any of the households in the population. Here there is no "aggregation problem."

Now let's complicate the example slightly, by supposing that there are two types of households, say "small" and "large". Suppose further that the only behavioral differences between these households involves a minimum (subsistence) demand for the good; namely, small households have demand function

$$(2.3) \qquad y_{it} = a_0(p_t) + b(p_t)M_{it}, \quad \text{family } i \text{ small}$$

and large families have demand function

$$(2.4) \qquad y_{it} = a_1(p_t) + b(p_t)M_{it}, \quad \text{family } i \text{ large},$$

where $a_0(p_t)$ and $a_1(p_t)$ represent the subsistence level demands. These forms of demand would arise from quasi-homothetic preferences for each type of family.

Suppose further that there are N_{0t} small families and N_{1t} large families, and $P_{0t} = N_{0t}/N_t$, $N_{1t}/N_t = 1 - P_{0t}$ denote percentages of small and large families. Aggregate demand is given as

$$(2.5) \qquad \bar{y}_t = a_1(p_t) + [a_0(p_t) - a_1(p_t)]\, P_{0t} + b(p_t)\, \bar{M}_t \,.$$

The impact of an additive difference among households is to introduce the percentage breakdown of household types (P_{0t}) into the aggregate equation. The response to a change in aggregate income $\bar{M}_t$ remains interpretable and well defined: a change in incomes causes a marginal adjustment for every family in line with $b(p_t)$, which matches the aggregate "effect" $b(p_t)$. If the population remained stable over time; with $P_{0t} = P_0$, then econometric estimation based on the equation

$$(2.6) \qquad \bar{y}_t = \tilde{a}(p_t) + \tilde{b}(p_t)\, \bar{M}_t$$

will uncover the marginal response $\tilde{b}(p_t) = b(p_t)$ and the average minimum demand $\tilde{a}(p_t) = a_1(p_t) + [a_0(p_t) - a_1(p_t)]\, P_0$.

However, if the population is not stable, with P_{0t} time varying, then econometric analysis based on equation (2.6) would generally not uncover the true income effect. If the percentage of small households trended with average income; namely up to error we have $P_{0t} \cong \vartheta + \kappa \bar{M}_t$, then the estimated effect of average income would be approximately $\tilde{b}(p) \cong b(p_t) + [a_0(p_t) - a_1(p_t)] \kappa$. At any rate, a correct specification requires including the composition effect P_{0t} into the aggregate equation, to permit measurement of the correct income effect. Of course, if the trend effect κ were minor, $\tilde{b}(p_t)$ would roughly measure $b(p_t)$, the marginal income response of each individual family.

In this example, the primary issue involves separation of the (well-defined) aggregate income effect $b(p_t)$ from the composition effect, which is accomplished by including the percentage P_{0t} in the aggregate demand equation. With any other kind of individual heterogeneity, this simple kind of separation is obliterated. In particular, one immediately faces the question of what the "aggregate income effect" is, or what "effect" would be measured by an econometric analysis of aggregate data alone.

In particular, now suppose that large and small households have different marginal responses to income; namely small households have demand function

(2.7) $$y_{it} = b_0(p_t) M_{it}, \quad \text{family } i \text{ small}$$

and large households have demand

(2.8) $$y_{it} = b_1(p_t) M_{it}, \quad \text{family } i \text{ large},$$

where we have omitted the subsistence levels ($a(p_t)$'s) for simplicity. These demands arise if all small households have identical homothetic preferences, as do all large households, but that preferences differ between small and large. In this case, aggregate demand is given as

$$\bar{y}_t = b_0(p_t)\, P_{0t}\, \bar{M}_{0t} + b_1(p_t)\,(1-P_{0t})\, \bar{M}_{1t} \,, \tag{2.9}$$

where

$$\bar{M}_{0t} = N_{0t}^{-1} \sum_{i \text{ "small"}} M_{it} \tag{2.10}$$

$$\bar{M}_{1t} = N_{1t}^{-1} \sum_{i \text{ "large"}} M_{it} \tag{2.11}$$

denote average income for "small" and "large" households respectively. Equation (2.9) reflects the fact that it now matters who gets the additional income, as small households spend it differently from large households. A correct implementation of this model would involve estimating equation (2.9), employing data on p_t, P_{0t}, $\bar{M}_{0t}$ and $\bar{M}_{1t}$.

However, out of practical expediency, suppose one fit

$$\bar{y}_t \cong B(p_t)\, \bar{M}_t \tag{2.12}$$

"as a good approximation", regarding $B(p_t)$ as a sort of "average" effect. To judge this approach, consider rewriting the true model (2.9) in terms of a "typical" income effect. While we could define this effect in various ways, let's take the most natural, namely the average income effect across all families:

$$\bar{b}(p_t) = P_{0t}\, b_0(p_t) + (1-P_{0t})\, b_1(p_t) \,. \tag{2.13}$$

With this assignment, we can rewrite the true equation (2.9) as

$$\bar{y}_t = \bar{b}(p_t)\, \bar{M}_t + D_t \tag{2.14}$$

which gives the "typical" aggregate effect, plus a distributional term

$$(2.15) \qquad D_t = [b_1(p_t) - b_0(p_t)]\, P_{0t}(1-P_{0t})\, [\bar{M}_{1t} - \bar{M}_{0t}] \quad .$$

This term depends on the difference in marginal effects $(b_1(p_t) - b_0(p_t))$, the composition of the population P_{0t}, and the relative distribution of income $(\bar{M}_{1t} - \bar{M}_{0t})$ over large and small families.

Several ideas come to mind for justifying the approximate equation (2.12) for estimation with aggregate data; we now take them in turn to keep the issues in focus. First, we note that if the percentage of small families P_{0t} varies over time t, then the typical effect $\bar{b}(p_t)$ of (2.13) likewise varies over time, so that the estimated aggregate coefficient $B(p_t)$ would attempt to measure a moving target. Let's assume this away by supposing that the population composition is indeed fixed with $P_{0t} = P_0$, or is so stable that this is a good approximation.

This pins down the typical effect $\bar{b}(p_t)$, so we now turn to the impact of omitting the term D_t. There is heterogeneity in marginal responses; $b_1(p_t) - b_0(p_t) \neq 0$; so we focus on the relative income term $\bar{M}_{1t} - \bar{M}_{0t}$. Suppose this difference trends with mean income; as in

$$(2.16) \qquad \bar{M}_{1t} - \bar{M}_{0t} \cong \tau\, \bar{M}_t \; .$$

Trending such as in (2.16) emerges if the distribution of income across families is constant, with $\bar{M}_{0t}/\bar{M}_t$ and $\bar{M}_{1t}/\bar{M}_t$ constant over time. In this case estimating equation (2.12) would give

$$(2.17) \qquad B(p_t) = \bar{b}(p_t) + [b_1(p_t) - b_0(p_t)]\, P_0(1-P_0)\, \tau \quad .$$

so that the macro coefficient $B(p_t)$ is a stable but biased measure of the typical effect $\bar{b}(p_t)$. We could, of course, beg the question by redefining the "typical" effect to equal the expression for $B(p_t)$ in (2.17). While this is silly, we return to this point later.

The mismeasurement caused by ignoring distributional composition is eliminated if the term D_t itself is negligible. We assume this by taking $\bar{M}_{0t} \cong \bar{M}_{1t}$, which gives $\tau \cong 0$ in (2.17). This assumption implies a zero correlation between incomes and the marginal income effects $b_0(p_t)$, $b_1(p_t)$ in each time period, and gives the true aggregate relation as

$$(2.18) \qquad \bar{y}_t = \bar{b}(p_t)\,\bar{M}_t \quad .$$

As such, estimating equation (2.12) would give $B(p_t) = \bar{b}(p_t)$.

It is clear under these assumptions that (2.18) is the true model, and that an econometric approach based on (2.12) would reveal this relationship. Because implementing (2.12) involves no omitted terms, specification tests could not reject (2.12), so that we would have empirical confirmation of this aggregate model. Indeed, the aggregation problem is solved, if a well fitting aggregate model is the overall goal. In fact, that same goal was attained under (2.16) when $\tau \neq 0$, where the aggregate coefficient $B(p_t)$ was given by (2.17). But the $\tau = 0$ case underlying (2.18) gives foundation for the interpretation that $B(p_t)$ is a "typical" income effect, namely that $B(p_t)$ as the average income effect $\bar{b}(p_t)$. As such, we have shown how to "solve" the aggregation problem in our example.

Or have we? Equation (2.18) is a well-specified, interpretable model that represents the aggregate data pattern without systematic error. But is that equation useful or valuable to an application? Would (2.18) adequately track the impact of changes in income on demand, under our behavioral model (2.7), (2.8)?

To pose this question, suppose that we wished to predict the impact of an increase in average income of $\Delta\bar{M} = 1$ at price level p_t. Some of the additional income will go to small families, say a fraction δ, with the rest $(1 - \delta)$ going to large families. This would entail a change in aggregate demand of

$$(2.19) \qquad \Delta\bar{y} = \delta\, b_0(p_t) + (1 - \delta)\, b_1(p_t) \ ,$$

reflecting the marginal spending responses of small and large families. On the other hand, equation (2.18) predicts the impact as

$$(2.19) \qquad \Delta\bar{y} = \bar{b}(p_t)\ \Delta\bar{M} = \bar{b}(p_t)$$

or the average income effect. Therefore, equation (2.18) gives an accurate prediction only when $\delta \cong P_0$, or when the additional income is distributed in a way that is uncorrelated with the marginal effects $b_0(p_t)$, $b_1(p_t)$. Therefore, for (2.18) to be accurate, any predicted income changes have to have the same distributional structure as assumed for justifying the equation to begin with. The same is true under (2.16) with $\tau \neq 0$; the "aggregate effect" (2.17) will accurately predict the impact of changing average income only when the new income is distributed in a fashion that maintains (2.16). In sum, while the above assumptions produce an equation that exactly fits existing data patterns, every one of those assumptions must hold for the estimated equation to have any practical value, including the assumptions on purely distributional features of the population. Neglecting distributional features undercuts the foundation of any equation based entirely on aggregate variables.

The "aggregation problem" is simply stated. Any incomplete summary of heterogeneous behavioral reactions, such as a relationship among aggregates, will fail in systematic ways to take account of those behavioral reactions. The "solution" is likewise obvious, namely that models need to account for heterogeneity and the composition of the population explicitly. The real issue in the last example above is that the true model is given by equation (2.9), which captures heterogeneity in marginal responses as well as the relevant distributional structure. Any simplification down to simple averages misses structure inherent to the basic behavioral reactions, which in turn, severely limits the usefulness of the simplified model.

Models that account for individual heterogeneity will typically not be estimable using data on economy-wide averages alone; additional data on distributional composition (such as P_{0t}, $\bar{M}_{0t}$, $\bar{M}_{1t}$ above), or micro data on individual behavior, will need to be incorporated. This should come as no surprise; to study relations that involve heterogeneous individual responses without distributional information is analogous to studying dynamic relations without using data over time. Moreover, with a properly specified model the incorporation is not difficult: the fact that a model ascribes structure to individual behavioral reactions implies that it is applicable in a consistent fashion to individual as well as aggregate data. The structure of individual responses, as well as necessary distributional assumptions, become an integral part of a properly specified model of aggregate data, and can provide testable restrictions that cannot be detected with aggregate data alone. Our survey discusses recent methods of econometric modeling that introduce these kinds of structure.

Most of the modeling methods involve fairly simple, sometimes static models of individual behavior. In contrast, the "representative agent" approach has been the vehicle for the development of fairly complex nonlinear models of individual behavior under uncertainty, and one might rightfully question whether our simple static examples above are not too simple, making more of heterogeneity issues than other familiar problems. In this regard, two observations are warranted. First, issues of individual heterogeneity are intrinsic to the use of aggregate data, whether individual models are static or dynamic. There is nothing in the economics of decision making over time or equilibrium theory which alters that fact, and the issues of heterogeneity and interpretation are worse for complicated nonlinear individual models than for simpler ones. There is simply no reason for according the "aggregation problem" a secondary status relative to other concerns (aside from ill-advised modeling convenience), as in representative agent modeling. Second, part of our survey will be to discuss some interesting interplay between the problem of aggregation and observed dynamic structure of aggregate data. One type of work shows how the failure to account for individual heterogeneity in an aggregate equation, which amounts to an omission of distributional effects, leads to spurious evidence of dynamic structure. Another type of work shows how aggregation over individual time series processes leads to more complicated dynamic structure among aggregate variables. Consequently, empirical issues of individual heterogeneity and dynamic structure in aggregate data are intertwined, with the assessment of their relative empirical importance yet to be settled.

We separate our discussion into two parts; theoretical modeling considerations in Section 3 and specific empirical models in Section 4. While Section 3 contains the principles that guide our discussion of specific models, this section can be read separately. Section 3.5 covers some broad issues of estimation, which are applicable to estimation of the empirical models of Section 4.

3. Theoretical and Econometric Considerations in Accounting for Aggregation over Individuals

Every model that accounts for aggregation over individuals must begin with a specification of individual behavior, or an econometric model applicable to individual level data. With regard to studying aggregate demand, as in section 2, the first step is to model the individual demand functions $y_{it} = f_i(p_t, M_{it})$, for each individual agent. In turn, this requires identifying the individual attributes that affect individual demands, including observable differences and differences that are modeled stochastically. We summarize the differences compactly as A_{it}, and rewrite the (common) individual demand function as $y_{it} = f(p_t, M_{it}, A_{it})$.

We use this simple demand paradigm to lay out the basic issues below, but there is nothing that restricts our treatment to the names given these variables above. The framework and the issues to be discussed below are applicable quite generally, to static and dynamic empirical models, and not just to demand models, as our terminology might suggest. There is no substantive difference between M_{it} and A_{it} as regards aggregation - both vary over individuals - and we keep them separate only to focus on a specific economic aggregate of interest, namely $\bar{M}_t$. The generic role of the price argument p_t is to represent variables common to all individuals, which do not introduce heterogeneity by themselves. The essential feature of the framework is the delineation of aspects that vary over individuals and aspects that are common across individuals.[3]

The "model" for aggregate demand $\bar{y}_t$ then appears simply as

$$\bar{y}_t = \frac{1}{N_t} \sum_i f(p_t, M_{it}, A_{it}). \tag{3.1}$$

If the population size N_t is large enough to appeal to a statistical law of large numbers, then we can associate $\bar{y}_t$ with the mean $E_t(y)$, using the formulation

$$E_t(y) = \int f(p_t, M, A)\, d\Omega_t(M, A) \tag{3.2}$$

where Ω_t is the distribution of M,A at time t. This formulation is generally necessary when (statistical) regularities are assumed for the distribution of individual variables.

The approaches we discuss involve different ways of implementing (3.1) or (3.2) in terms of modeling aggregate data. Exact aggregation and related linear methods are based on restrictions on the form of f(.) to structure (3.1) or (3.2). Nonlinear individual models with distribution restrictions, or restrictions on the structure of Ω_t, give another way of implementing (3.2). Finally, a further possibility is to characterize the individual function $f(p_t, M_{it}, A_{it})$ completely, with cross-section and/or panel data on individuals. The "micro-simulation" approach predicts $\bar{y}_t$ by implementing (3.1) (or (3.2)) directly, by explicit addition over agents, with Ω_t the observed empirical distribution at time t.

3.1 The Role of Linearity and Exact Aggregation

The exact aggregation approach involves restricting the model for individual behavior so as to limit the amount of distributional information required for the implied aggregate model.[4] Eliminating the need for distributional distinctions often requires fairly strong linearity restrictions on the individual model. The theory underlying exact aggregation methods is often couched in overly strong terms of when a generic form of aggregate equation "exists," which just reflects the idea that if distributional effects belong in a model, then a model without such effects doesn't "exist."

The essence of exact aggregation theory can be seen from the original question of the foundation of a per-capita demand equation for a commodity, or the simplest form of representative agent model. In particular, when are we permitted to model average demand $\bar{y}_t$ as a function of average income $\bar{M}_t$ and prices p_t? More formally, when can we assert that

(3.3) $$\bar{y}_t = F(p_t, \bar{M}_t) \quad ;$$

without attention to individual heterogeneity?

The basic logic of what equation (3.3) says is sufficient to ascertain its implications. As long as average income $\bar{M}_t$ (or p_t) does not change, then neither does $\bar{y}_t$. Consider what this means. Suppose you were to reach into my pocket, and take a fifty dollar bill. I would be poorer and you richer, and both of us would adjust our purchases of the commodity in question. Because average income $\bar{M}_t$ has not changed, equation (3.3) implies that average demand $\bar{y}_t$ does not change, which means that my purchase adjustment must be exactly offset by yours. In other words, our marginal reactions to a change in income *must* coincide. However, equation (3.3) is not affected by how much money is taken, or whose pockets are involved in such transfers, so we must conclude that everyone's marginal reactions are the same. Individual demands must be of the form

(3.4) $$f(p_t, M_{it}, A_{it}) = a(p_t, A_{it}) + b(p_t)\, M_{it},$$

or that individual Engel curves are parallel and linear. This gives the aggregate demand function as

(3.5) $$F(p_t, \bar{M}_t) = N_t^{-1} \sum a(p_t, A_{it}) + b(p_t)\, \bar{M}_t \quad .$$

The aggregate income effect $b(p_t)$ is quite interpretable - it is the marginal income effect displayed by every individual in the population. To the extent that the population changes over time, or that equation (3.3) holds when the distribution of attributes $\{A_{it}\}$ is freely varied, then the logic extends to the intercept, giving

$$(3.6) \qquad f(p_t, M_{it}, A_{it}) = a(p_t) + b(p_t)\, M_{it},$$

so that no individual differences are allowed at all. If further, demand is zero when income is zero, then $a(p_t) = 0$ as well, with demand proportional to income for each family, and aggregate demand proportional to aggregate income.

The severity of these restrictions on individual behavior (no heterogeneity in marginal reactions) reflect the strength of the requirement of (3.3) that distributional effects are irrelevant. The exact aggregation approach is based on applying the logic above in weakened form, with distributional elements introduced in a controlled fashion. To set ideas, recall the example of Section 2 above where "small" and "large" families displayed different propensities to consume. In our present notation, let the attribute vector A_{it} be a qualitative variable, with $A_{it} = 1$ denoting a small family and $A_{it} = 0$ denoting a large family. The basic model (2.7) and (2.8) is compactly written as

$$(3.7) \qquad y_{it} = b_0(p_t)\, A_{it}M_{it} + b_1(p_t)\,(1-A_{it})M_{it}\ .$$

$$= b_1(p_t)\, M_{it} + [b_0(p_t) - b_1(p_t)]\, A_{it}M_{it}$$

The model for aggregate demand (2.9) is then written as

$$(3.8) \qquad \bar{y}_t = b_1(p_t)\, \bar{M}_t + [b_0(p_t) - b_1(p_t)]\, \overline{AM}_t$$

where

$$(3.9) \qquad \overline{AM}_t = N_t^{-1} \sum A_{it}M_{it} = N_t^{-1} \sum_{i \text{ "small"}} M_{it}$$

This matches (2.9), as $\overline{AM}_t = P_{0t}\bar{M}_{0t}$, where $P_{0t} = N_t^{-1} \sum A_{it} = \bar{A}_t$. Here, the form of the individual demand function establishes what distributional information is required, namely $\overline{AM}_t$, as well as how to interpret the macro coefficients. In particular, the coefficient of $\bar{M}_t$ is the marginal propensity to consume $b_1(p_t)$ of "large" families, and the coefficient of $\overline{AM}_t$ is the difference $b_0(p_t) - b_1(p_t)$ of the propensity to consume between "small" and "large" families.

The theory of exact aggregation focuses on the aggregate equation, insisting that it depend on only a small number of distributional statistics. In particular, one can ask what restrictions are implied if the aggregate equation takes the form.

$$(3.10) \qquad \bar{y}_t = F(p_t, \mathcal{M}_{1t}, \mathcal{M}_{2t}, \ldots, \mathcal{M}_{Jt})$$

where the $\mathcal{M}_{jt}$ arguments are J statistics of the joint income-attribute $\{(M_{it}, A_{it})\}$ distribution;

$$(3.11) \quad \mathcal{M}_{jt} = \mathcal{M}_j[(M_{1t}A_{1t}), (M_{2t}, A_{2t}), \ldots, (M_{N_t t}, A_{N_t t})] \quad j = 1, \ldots, J \quad .$$

This generalizes (3.3), in which J = 1 and $\mathcal{M}_{1t} = \bar{M}_t$. As in the more restricted problem, the ability to vary the joint income-attribute distribution enforces intrinsic linearity on the individual demand equation, as well as requiring the distributional statistics ($\mathcal{M}_j$'s) to be averages. A precise version of this theory is given in Jorgenson, Lau and Stoker(1982), Lau (1977, 1982) and others. The argument is given loosely as follows.

The main requirements of this theory are that the distributional statistics are not redundant among themselves or in aggregate demand, and that the joint income-attribute distribution can be varied arbitrarily. The first feature is used to establish that the distributional statistics $\mathcal{M}_{1t}, \ldots, \mathcal{M}_{Jt}$ are functions of averages, and so can be taken as averages themselves.

Therefore, the permissible distributional statistics are sample moments; say with

$$(3.12) \qquad \mathcal{M}_{jt} = N_t^{-1} \sum x_j(M_{it}, A_{it}), \qquad j = 1, \ldots, J,$$

where $x_j(M,A)$ is a function of the individual income and attribute values.

The second feature, arbitrary variation of the distribution, is then applied to show that the individual demands must be intrinsically linear. The conclusion of this argument shows that the $x_j(M,A)$ can be redefined so that a marginal change in $x_j(M,A)$ depends only on p, with individual demand taking the linear form

$$(3.13) \qquad f(p_t, M_{it}, A_{it}) = a(p_t) + b_1(p_t)x_1(M_{it}, A_{it}) + \ldots + b_J(p_t)x_J(M_{it}, A_{it})$$

$$= a(p_t) + b(p_t)^T x(M_{it}, A_{it}) \quad ;$$

where $b(p) \equiv (b_1(p), \ldots, b_J(p))'$ and $x(M,A) \equiv (x_1(M,A), \ldots, x_J(M,A))$. This, in turn, gives that aggregate demand is linear in the sample moments, namely that

$$(3.14) \qquad \bar{y}_t = a(p_t) + b_1(p_t)\ [N_t^{-1} \sum x_1(M_{it}, A_{it})]$$

$$+ b_2(p_t)[N_t^{-1} \sum x_2(M_{it}, A_{it})] + \ldots + b_J(p_t)[N_t^{-1} \sum x_J(M_{it}, A_{it})]$$

$$= a(p_t) + b(p_t)'\bar{x}_t$$

with $\bar{x}_t \equiv N_t^{-1} \sum x(M_{it}, A_{it})$. This type of generalized linear structure for both micro and macro level equations is the characterizing feature of exact aggregation models. Again, it is important to stress how this structure applies generally to aggregation problems (and is in no way restricted to demand analysis).

Within the context of demand analysis, the components of $x(M,A)$ can represent linear and nonlinear functions of income, as well as functions of observable differences across families. The model (3.7) has $x_1(M,A) = M$ and $x_2(M,A) = MA$, and we consider more extensive exact aggregation models below. It is important to note that nonlinear terms in income M likewise give rise to marginal differences as above; for instance various demand models take the form

$$(3.15) \qquad y_{it} = b_0(p_t) M_{it} + b_1^*(p_t) A_{it}M_{it} + b_2^*(p_t) M_{it}\ln M_{it}$$

which leads to an entropy measure $N_t^{-1} \sum M_{it}\ln M_{it}$ in the equation for aggregate demand $\bar{y}_t$. Further, much of the work on exact aggregation demand models also uses economic optimization theory to structure the income effects; for instance, the budget constraint ("adding-up") of demands implies that a system in exact aggregation form must have $x_1(M,A) = M$, and homogeneity of degree zero in prices and income likewise restricts the form of further income terms and the coefficients (as functions of prices). At any rate, specific models usually reflect restrictions to deal with aggregation, as well as restrictions from the underlying individual optimization theory.

The practical attractiveness of exact aggregation models derives from three sources. First, the aggregate equations can be immediately derived from the individual equations, with the distributional impacts clearly interpretable. In particular, having specified an individual demand equation of the form (3.13), the aggregate equation can immediately be written down, and the required distributional statistics ($\bar{x}_t$) stated. This practical ease in modeling cannot be overstated. Second, while intrinsic linearity may appear as a stringent requirement, the fact that virtually any function of individual attributes can be used permits a wide range of heterogeneous responses to be modeled - any area using linear models for survey or panel data analysis has exploited such restrictions. Moreover, any specific set of equation restrictions can be tested statistically with data on differing individuals, either from a cross section survey or a panel survey.

Third, and perhaps most important, is that exact aggregation models are fully interpretable. The individual level model is fully recoverable from the aggregate model, because the coefficient functions of $y_{it} = a(p_t) + b(p_t)^T x(M_{it}, A_{it})$ match those of the aggregate model $\bar{y}_t = a(p_t) + b(p_t)^T \bar{x}_t$. While obvious, it is important to recall what this means for the use of economic theory to restrict aggregate models. For modeling demand, the individual coefficient functions are structured by integrability conditions, and the same restrictions are applicable to the aggregate data model. This does not mean that the aggregate demand equations are integrable themselves, but just that the full modeling benefits of rational individual choice are available for the aggregate model.

3.2 Nonlinearity, Distributional Restrictions and Recoverability

While exact aggregation models are applicable in a variety of areas, there are settings where the intrinsic linearity of such models is unwarranted or undesirable. When individual behavioral relations are nonlinear, then exact aggregation theory is not applicable, and so one might ask how a model could be built that accounts for aggregation over individuals. Permitting arbitrary variations in the underlying distribution of individual attributes brought about micro linearity, so this feature must be dropped. In particular, the structure of the distribution of individual attributes must be included as part of a model that accounts for heterogeneity of individual responses. This change of posture also requires some rethinking of the basic issues surrounding interpretability of the relationship between aggregates.

As before, the issues are best illustrated with an example. Suppose that we are studying the purchase of a single unit of a particular product, and we only observe whether it is bought (say $y_{it} = 1$) or not ($y_{it} = 0$). Assume for the moment that the value to family i of buying this product depends only on the price p_t of the product and the family's overall budget M_{it}; in particular, suppose the net benefits (utility) are modeled as $1 + \beta_1 \ln p_t + \beta_2 \ln M_{it}$. The individual model of purchase is then

$$(3.16) \qquad y_{it} = f(p_t, M_{it}) = 1 \quad \text{if } 1 + \beta_1 \ln p_t + \beta_2 \ln M_{it} \geq 0$$

$$= 0 \qquad \text{otherwise}$$

Because of the (0,1) nature of y_{it}, this model is nonlinear in $\ln M_{it}$, and cannot be made to be linear in a function of $\ln M_{it}$ or M_{it} that does not depend on the parameters β_1, β_2 (or be put in exact aggregation form). Indeed, addition of a normal error term on the right-hand-side of (3.16) would give a probit model.

The aggregate $\bar{y}_t = N_t^{-1} \sum y_{it}$ here is the proportion of all families that buy the product. How is this proportion to be modeled in a manner consistent with the individual model, at least for a large population?

For this it is necessary to structure the distribution of M_{it}, and derive the aggregate model as the probability that a purchase is made. With the distribution restriction, the aggregate model is derived in a straightforward fashion from (3.2). Consequently, we suppose that the distribution of M_{it} is lognormal in each period t, say with $\ln M_{it}$ having mean μ_t and variance Σ_t^2.

To derive the aggregate relation, consider the probability $E_t(y)$ of purchase, or the probability of

$$(3.17) \qquad 1 + \beta_1 \ln p_t + \beta_2 \ln M \geq 0 \quad .$$

Some arithmetic[5] gives this as the probability of

$$(3.18) \qquad -\Sigma_t^{-1}(\ln M - \mu_t) \leq \frac{1}{\beta_2 \Sigma_t}\left(1 + \beta_1 \ln p_t + \beta_2 \mu_t\right)$$

where the left-hand variable is normally distributed with mean 0 and variance 1. Therefore, we have that

$$(3.19) \qquad E_t(y) = \Phi\left[\frac{1}{\beta_2 \Sigma_t}\left(1 + \beta_1 \ln p_t + \beta_2 \mu_t\right)\right]$$

where $E_t(y)$ is the fraction of families purchasing the product, $\Phi(.)$ is the univariate normal cumulative distribution function and μ_t is the mean of log income. To rewrite this equation in terms of mean income $E_t(M)$, we again appeal to the lognormal assumption, for which $E_t(M) = \exp\,[\mu_t + (1/2)\Sigma_t^2]$. Solving this for μ_t and substituting into (3.19) gives the aggregate model as

$$(3.20) \qquad E_t(y) = \Phi\left[\frac{1}{\beta_2 \Sigma_t}\left(1 + \beta_1 \ln p_t + \beta_2 \ln E_t(M) - \beta_2 \frac{\Sigma_t^2}{2}\right)\right]$$

Thus, the proportion of families buying the product is a nonlinear (cummulative normal) function of the product's price and of the mean and (log) variance of family income. With observations on $E_t(y)$, $E_t(M)$ (or $\bar{y}_t$, $\bar{M}_t$) and Σ_t over time t, the (individual level) behavioral parameters β_1 and β_2 could be estimated. Note that if Σ_t were constant (say Σ) over time, then it could be estimated as a parameter as well.

The impact of heterogeneity in (3.20) is most evident because of the appearance of Σ_t, gauging the spread of (log) income. Also of interest is the appearance of something we might call an "aggregate net benefit of purchase," namely $1 + \beta_1 \ln p_t + \beta_2 \ln E_t(M)$. While one might find it convenient to name this expression in this fashion, it is clear that such a "net benefit" has no behavioral interpretation; no "agent" formulates a decision on the basis of it. The model of purchase choice is at the individual level, where it needs to be to give foundation to the interpretation of β_1 and β_2.

If other elements of individual heterogeneity are relevant to this purchase decision, then more distributional information is necessary in the aggregate model. For instance, suppose that the net benefits differ between "small" families ($A_{it} = 1$) and "large" families ($A_{it} = 0$), in accordance with the model

$$(3.21) \qquad y_{it} = f(p_t, M_{it}) = 1 \quad \text{if } 1 + \beta_1 \ln p_t + \beta_2 \ln M_{it} + \beta_3 A_{it} \geq 0$$

$$= 0 \qquad \text{otherwise} \quad .$$

We now structure the joint distribution of M_{it} and A_{it} to model the overall probability of buying $E_t(y)$. Denote the proportion (probability) of small families as $P_{0t} = E_t(A)$, and assume that the income of small families is lognormally distributed with mean $E_{0t}(M)$ and log-variance Σ_{0t}^2, and that the income of large families is lognormally distributed with mean $E_{1t}(M)$ and log-variance Σ_{1t}^2. The aggregate model now is

$$(3.22) \qquad E_t(y) = P_{0t} \, E_t(y|A = 1) + (1-P_{0t}) \, E_t(y|A = 0)$$

$$= P_{0t}\, \Phi\left[\frac{1}{\beta_2 \Sigma_{0t}}\left(1 + \beta_1 \ln p_t + \beta_2 \ln E_{0t}(M) - \beta_2 \frac{\Sigma_{0t}^2}{2} + \beta_3\right)\right]$$

$$(1 - P_{0t})\, \Phi\left[\frac{1}{\beta_2 \Sigma_{1t}}\left(1 + \beta_1 \ln p_t + \beta_2 \ln E_{1t}(M) - \beta_2 \frac{\Sigma_{1t}^2}{2}\right)\right]$$

While a more complicated equation, the same features are retained, namely the individual model parameters β_1, β_2 and β_3 could be estimated with aggregate data (including the distributional variables). Here, this model has nothing to do with an "aggregate net benefit" $1 + \beta_1 \ln p_t + \beta_2 \ln E_t(M) + \beta_3 P_{0t}$, not that any such connection would ever be expected.

These examples point out how aggregate models can be formulated with nonlinear individual models. Also, they stress the importance of interpreting the model parameters in terms of the original, individual level, model. We assumed specific forms for the distributions of underlying attributes - these features are a necessary part of the model, and could be tested, as with any other feature of model specification.

Because of these features, it is natural to think that the use of distributional restrictions would eliminate all of the problems posed by aggregation over individuals. In one sense this is true, but in another it is not. In particular, the foundation of the aggregate model rests on its connection to individual behavior, in that the behavioral parameters are recoverable from the aggregate model. While an aggregate relationship can always be characterized statistically, at least in principle, it is not interpretable, nor can it be counted on to track the aggregates out of the statistical sample. Without such a recoverability property, there is no clear connection between aggregate data patterns and individual behavior.

The basic recoverability issue is fairly easy to spell out. We alter our notation slightly, denoting the individual model as $y_{it} = f(p_t, x_{it}, \beta)$, where x_{it} summarizes individual attributes (or functions of observed attributes such as M, A above), and β represents parameters of interest. Suppose that the distribution of x at time t is given as $\Omega_t(x) = \Omega(x, \mu_t)$, where we have used parameters μ_t (say $\mu_t = (E_t(x), \Sigma_t, \ldots)$) to summarize how the distribution varies over time t. The aggregate relation is given from (3.2) as

$$(3.23) \qquad E_t(y) = \phi(p_t, \mu_t, \beta) = \int f(p_t, x, \beta)\, d\Omega(x, \mu_t) \quad .$$

Individual behavior is recoverable from this aggregate relation if β is identified by the formula (3.23). This occurs if ϕ always changes when β is varied (regardless of how β is varied), or in other words, given a sufficient number of observations on $E_t(y)$, p_t, μ_t that fit equation (3.23), it is possible to solve for β uniquely. This was true for the examples above, but it need not be true for any specification of f and/or Ω.

This issue is studied in some detail in Stoker (1984a), where the focus is on heterogeneity per se, or with the argument p_t held constant. Some of the results of this analysis are of interest here. First, one can verify that linear individual models are the only models that give recoverability for broad ranges of distributions, which is a verification of exact aggregation theory in the large sample context. Second, there are classes of distributional restrictions where recoverability is assured regardless of the form of the individual behavioral model. These classes are known as "complete" distribution classes in statistics, with the foremost example being distributions of the exponential family. This family refers broadly to a distribution restriction of the form

(3.24) $$d\Omega(x,\mu_t) = \rho_0(x)\ c(\mu_t)\ e^{\pi(\mu_t)'D(x)}\ dx,$$

or where a base density $\rho_0(x)$ is altered over time in a fashion consistent with the exponential term above (permitting unconstrained variations in $\pi(\mu_t)$). The exponential family contains several familiar distributional formulations; normal, gamma, beta, as well as the lognormal distribution used above. In these cases recoverability is assured, and estimation of all the behavioral parameters can be based on aggregate data alone.[6]

3.3 Differences Between Aggregate and Individual Models

Situations where recoverability fails often provide the key to understanding differences between models estimated with individual and aggregate level data. Such situations arise because the distribution of heterogeneous attributes fails to vary sufficiently for the effects of the attributes to be measured. Perhaps the clearest way to see this point is to consider restrictions associated with the sort of "aggregation factors" used by Arthur Lewbel (1991) and Richard Blundell, Panos Pashardes and Guglielmo Weber (1992) (to be discussed later). Suppose that the individual attribute variables are partitioned as $x = (x_1,x_2)$, x_1 a single variable, and the individual model is in exact aggregation form; say[7]

(3.25) $$y_{it} = a(p_t) + b_1(p_t)'x_{1it} + b_2(p_t)'x_{2it} \quad ,$$

so that the correct aggregate model is

(3.26) $$E_t(y) = a(p_t) + b_1(p_t)'E_t(x_1) + b_2(p_t)'E_t(x_2) \quad .$$

Consider a couple kinds of constancy restrictions, namely i) $E_t(x_2) = c_1$, constant, or ii) $E_t(x_2)/E_t(x_1) = c_2$, constant. In both cases the aggregate relationship is linear in $E_t(x_1)$, as

$$(3.27) \qquad E_t(y) = \tilde{a}(p_t) + \tilde{b}_1(p_t)'E_t(x_1),$$

However, the correspondence of (3.27) with the individual model (3.25) differs, depending on how recoverability fails. Specifically, under i) we have $\tilde{a}(p_t) = a(p_t) + b_2(p_t)'c_1$, $\tilde{b}_1(p_t) = b_1(p_t)$, and under ii) we have $\tilde{a}(p_t) = a(p_t)$, $\tilde{b}_1(p_t) = b_1(p_t) + b_2(p_t)'c_2$. In one instance $b_1(p_t)$ can be recovered, but not $a(p_t)$ or $b_2(p_t)$, and in the other case $a(p_t)$ can be recovered, but not $b_1(p_t)$ or $b_2(p_t)$. Nevertheless, if one could verify these kinds of constancy restrictions in a particular data set, one has an explanation for the aggregate model (3.27) together with the individual model (3.25). For instance, if the "aggregation factor" $E_t(x_2)/E_t(x_1)$ were constant, then (3.27) would be useful for prediction in situations where the factor remained constant.

The effects of certain individual attributes are impossible to measure with aggregate data when aspects of the heterogeneity distribution are strictly constant. Return to our general format, with $x = (x_1, x_2)$ as above. Suppose that given the value of x_1, the distribution of x_2 is constant over time. In other words, suppose that the density of the underlying distribution is structured as

$$(3.28) \qquad d\Omega(x,\mu_t)/dx = \rho_{2|1}(x_2|x_1)\ \rho_1(x_1|\mu_t) \ .$$

It is easy to see that this structure makes it impossible to study the effects of heterogeneity represented by x_2 with aggregate data. In particular,

$$(3.29) \qquad E_t(y) = \int f(p_t,x,\beta)\, d\Omega(x,\mu_t)$$

$$= \int\left[\int f(p_t,x,\beta)\, \rho_{2|1}(x_2|x_1)\, dx_2 | x_1\right]\rho_1(x_1|\mu_t)\, dx_1 \ .$$

$$= \int f^+(p_t,x_1,\beta)\, \rho_1(x_1|\mu_t)\, dx_1 \ .$$

where given x_1, f^+ is the mean value of f, or $f^+(p_t,x_1,\beta) = E[f(p_t,x,\beta)|x_1]$. In this setting, sufficient variation in the distribution of x_1 may permit recoverability of f^+. Recoverability of a more detailed individual model, such as f, is impossible, because there is variation only in the marginal distribution of x_1. From the vantage point of aggregate data, the empirical implications of beginning with the model f() are the same as beginning with the simplified model $f^+()$.

Recoverability can fail in many other ways, often resulting in an aggregate data pattern that has little resemblance to the individual behavioral model. One extreme case is where the underlying distribution just trends with the aggregates of interest. For instance, suppose $\mu_t = E_t(x)$, and that the density of the distribution is

$$(3.30) \qquad d\Omega(x,E_t(x))/dx = \rho_0(x) + [E_t(x)-E_0(x)]'s(x) \ .$$

Here $\rho_0(x)$ is a base density (say from one time period), and s(x) indicates how the density shifts with the aggregate $E_t(x)$; we have $\int \rho_0(x)dx = 1$, $\int x\rho_0(x)dx = E_0(x)$, $\int s(x)dx = 0$ and $\int x\, s(x)dx = (1,\ldots,1)'$. This structure says that any group of individuals defined by a fixed range of x, accounts for a proportion of the population that varies linearly with the mean $E_t(x)$. What affect would this have on aggregation? From (3.30), we have

$$(3.31)\quad E_t(y) = \int f(p_t,x_t,\beta)\ (\rho_0(x) + [E_t(x)-E_0(x)]\ s(x))\ dx$$

$$= \int f(p_t,x_t,\beta)\rho_0(x)dx \ + [E_t(x)-E_0(x)] \int f(p_t,x_t,\beta)\ s(x)\ dx$$

$$= a(p_t,\beta) + b(p_t,\beta)\ E_t(x) \quad ,$$

or the aggregate relationship is always linear in the mean $E_t(x)$. Regardless of whether the original model was highly nonlinear; say exponential, high degree polynomial, or even 0-1 as in the purchase example above, it is impossible to distinguish it from a linear individual model consistent with the above equation.[8] Of course it may be possible that particular choices of f, ρ_0 and s would result in β being identified by (3.31). But with distributional trending, the aggregate relation can bear little resemblance to individual behavior, with recoverability of any nonlinear individual model ruled out.

The cases where recoverability fails again point up that care is required in the applicability of restrictions from individual behavior to aggregate models. Each of the cases above involves too little independent variation in the population distribution over time to recover the individual model, which means that distribution effects exist but are not measurable with aggregate data alone. As such, these settings are ones in which simple aggregate data models will describe the data patterns exactly, but individual behavioral restrictions cannot be casually ascribed to such aggregate models.

It is important to keep in mind what these concerns are negative about. In particular, they are pessimistic regarding the prospects of learning about behavior from aggregate data alone. The solution is likewise simple; namely model individual behavior, use aggregate data in a fashion consistent with that individual model, and combine individual data in estimation when possible. If all one has is aggregate data, the recoverability property is essential for a model can be interpreted in terms of individual behavior. But for many (most) applications, there may be too much richness in individual behavior to expect that a few aggregate data series will reveal it adequately.

3.4 Unobserved Individual Heterogeneity and Stochastic Aggregation Models

A natural recourse for capturing the myriad of individual differences in many practical problems is to model such differences as unobserved random variables. In the context of models that deal with aggregation over individuals, one needs to pay special attention to how such unobserved attributes are distributed and how their distribution evolves over time. Moreover, various approaches to aggregation have unobserved individual differences as a starting point, and our discussion of random elements gives a natural format for discussing econometric estimation. For this discussion, we expand the notation so that the individual model is now

$$y = \tilde{f}(p, x, \beta, \epsilon) \tag{3.32}$$

where x (and its distribution) are observed and ϵ represents unobserved attributes, whose distribution must be modeled.

In the abstract, x and ϵ are indistinguishable, and so all of the above remarks about recoverability could apply for the recoverability of $\tilde{f}$ from the relation between aggregates. However, because ϵ and its distribution in any time period are not directly observed, we consider the situation where the density of (x,ϵ) factors as $\rho_\epsilon(\epsilon|x,\sigma)\ \rho(x|\mu_t)$; or that the density of ϵ for given x is stable at each time period, where we permit a vector of parameters σ.[9] The most straightforward setting for dealing with unobserved attributes is when their impact is additive, as in

$$(3.33) \qquad y = \tilde{f}(p,x,\beta,\epsilon) = f(p,x,\beta) + \epsilon$$

where we assume $E_t(\epsilon|x) = 0$, for each time period t. From exact aggregation theory, it is clear that (3.33) would be implied if the average of y depends in general on only the marginal distributions of x and ϵ.

The aggregate relationship is generally written as

$$(3.34) \qquad E_t(y) = \int\left(\int \tilde{f}(p_t,x,\beta,\epsilon)\ \rho_\epsilon(\epsilon|x,\sigma)\ d\epsilon|x\right)\rho(x|\mu_t)\ dx\ .$$

$$= \int f(p_t,x,\beta,\sigma)\ \rho(x|\mu_t)\ dx\ .$$

where $f(p,x,\beta,\sigma) = E[y|p,x]$. As this is a situation of conditional constancy, the conditional expectation $f(p,x,\beta,\sigma)$ captures all of the structure of the individual model for aggregate data. Recoverability would focus on whether the parameters β and σ could be identified with sufficient aggregate data.[10]

Two practical points are worth noting. First, the criterion for the inclusion of variables centers on the stability of the conditional expectation $E(y|x) = f(p,x,\beta,\sigma)$; omitting an important variable can cause this conditional expectation to vary over time. This is closely connected to the question of how behavioral regularities are ascribed across individuals - standard micro econometric models structure the effect of observed variables, but other approaches try to avoid this, summarizing differences through randomly varying parameters.

Second, with regard to the aggregate relation (3.34), there is no practical difference among any individual stochastic models with the same conditional expectation $E(y|x) = f(p,x,\beta,\sigma)$. That is, whether unobserved differences are in levels as in (3.33), or entered in a more complicated fashion, there is no effect on the mean aggregate relation. For instance, a random coefficient model

$$(3.35) \qquad y = \tilde{f}(x,\beta,\epsilon) = \beta_0 + (\beta_1 + \epsilon_1)'x + \epsilon_0$$

has the same aggregate empirical implications as a model with common coefficients (omitting ϵ_1) provided $E(\epsilon_1|x) = 0$, an observation due to Arnold Zellner (1969). This latter restriction is related to the familiar covariance restriction of linear aggregation, $Cov(x,\epsilon_1) = 0$ (which implies $E_t(y) = \beta_0 + \beta_1'E_t(x)$); in particular, $E(\epsilon_1|x) = 0$ is implied if the covariance restriction holds for all possible distributions of x. If in (3.35) the disturbances ϵ_0 are homoskedastic, then (3.35) implies increasing variance of y with increases in x, whereas a common coefficient model without ϵ_1 would have constant variance of y over x values. Of course, if there are coefficient patterns over different x values, or $E(\epsilon_1|x) = c(x) \neq 0$, then the appropriate, potentially recoverable regression exhibits those patterns, as in

$$(3.36) \qquad E(y|x) = f(x,\beta) = \beta_0 + [\beta_1 + E(\epsilon_1|x)]'x + E(\epsilon_0|x).$$

These notions illustrate the interplay between modeling individual differences and the observed variables. The most sensible posture is to use variables to represent all observable individual differences, interpreting the results of analysis the way one interprets a regression pattern estimated from individual level data.

3.5 Econometric Issues and Aggregation

Throughout this section we have discussed aggregation questions in the general context of recoverability of individual behavior from aggregate data patterns. In practical terms, we will typically have a micro model specified up to some parameter values, and the object of empirical work will be to estimate the parameters. It is necessary that such parameters be identified from all of the data available, including whatever individual data is relevant as well as aggregate data. Recoverability of certain parameters means that they are identified from the aggregate model alone.

Estimation of a well specified model that accounts for aggregation over individuals does not entail any nonstandard econometric issues. In particular, such a model involves estimation of a set of parameters over one or more data sources, and the only real concern is that the individual model is applied to data on individuals and the aggregate model is applied to aggregate data. Our purpose here is to complete our coverage by raising a few of the broad issues; namely to discuss estimation in the context of full or partial recoverability, as well as discuss some results that permit partial pooling of individual and aggregate data.

As above, we suppose that the individual model is denoted

(3.37) $$y = \tilde{f}(p_t, x, \beta, \epsilon)$$

where ϵ is random with density $\rho_\epsilon(\epsilon|x,\sigma)$, and the micro regression of y on x is denoted

(3.38) $$E(y|x) = f(p_t, x, \gamma) \ ,$$

where we denote all of the parameters for estimation as $\gamma = (\beta,\sigma)$. The aggregate model is given as

(3.39) $$E_t(y) = \int f(p_t, x, \gamma) \ \rho(x|\mu_t) \ dx \ = \phi(p_t, \mu_t, \gamma).$$

With full recoverability, when γ represents a small number of parameters relative to the number of aggregate observations, estimation can proceed on the basis of aggregate data alone. In particular, if $\bar{y}_t$, p_t, μ_t denoted the aggregate observations, then γ could be estimated consistently by (nonlinear) least squares, as $\hat{\gamma} = \text{argmin} \sum [\bar{y}_t - \phi(p_t, \mu_t, \gamma)]^2$. We could also consider weighted least squares estimators; James Powell and Stoker (1985) show how efficient weighting schemes can be implemented, using the stochastic structure imparted by aggregation across a random sample. In specific examples, it may be easy to derive a likelihood function for the aggregate estimation.[11]

When γ represents a large number of parameters measuring effects of many kinds of individual differences, there may not be sufficient aggregate data points either to identify all the components of γ, or to measure them with any precision. This situation is typical in realistic problems dealing with individual heterogeneity, and makes it necessary to bring more detailed data into the estimation process, such as data on individuals. If there is sufficient data at the individual level, such as a panel of many individuals over many time periods, there may be no inherent need to formulate a specific model for aggregates at all. For example, the parameters could be estimated by maximum likelihood methods, or if γ is identified in the regression (3.38), by (nonlinear) least squares regression.

While this is all quite standard, two remarks are called for. First, the only substantive reason for formulating an aggregate model when full panel data is available is to facilitate aggregate predictions - namely formalizing how the distribution of individual attributes varies in simulated time periods. Second, some standard panel data methods can seriously complicate attempts to model aggregates; for example, the incorporation of fixed individual effects. A fixed effects setup is only tractable when the number of individuals is relatively small, with aggregation carried out explicitly (say aggregation over counties in a state). Otherwise, the individual effects need to be regarded as random for overall aggregation, with their distribution (joint with observed individual variables) specified fully.[12]

These two situations represent two extremes, namely sole reliance on aggregate data versus sole reliance on individual data. In practical terms, the situations that fall between these two extremes are those that are best addressed with models that account for aggregation. That is, where the model involves sufficient numbers of individual differences to be realistic, but too many to be studied with aggregate data alone, and there is some limited individual level data, such as one or more cross section surveys.

This setting gives rise to measuring effects of individual heterogeneity with individual data, and measuring the effects of common variables with aggregate data. To outline how this is done, suppose that γ_{cs} represents the subvector of γ that is identified with cross section data at time t_0, that can be thought of as the parameters gauging the effects of individual differences. Suppose that γ_{ag} represents the subvector of γ that is identified in the aggregate data (namely in the model (3.39)), and where each element of γ appears in either γ_{cs}, γ_{ag} or both. In a demand modeling scenario, γ_{cs} could represent income and demographic effects, and γ_{ag} could represent price and income effects (through the impact of aggregate income). In this situation, γ_{cs} could be estimated by either maximizing period t_0 likelihood

(3.40) $$\hat{\gamma}_{cs} = \operatorname{argmax}_{\gamma_{cs}} \sum_i \ln \rho_y(y_{it_0}, p_{t_0}, x_{it_0}; \gamma)$$

where $\rho_y(\ )$ is the likelihood derived from the behavioral model (3.37) and the conditional distribution ρ_ϵ of ϵ given x. If γ_{cs} is identified by the regression (3.38), then least squares can be applied with the cross section data

(3.41) $$\hat{\gamma}_{cs} = \operatorname{argmin}_{\gamma_{cs}} \sum_i [y_{it_0} - f(p_{t_0}, x_{it_0}; \gamma)]^2 \quad .$$

The parameters γ_{ag} could then be estimated with the aggregate data via

(3.42) $$\hat{\gamma}_{ag} = \operatorname{argmin}_{\gamma_{ag}} \sum [\bar{y}_t - \phi(p_t, \mu_t, \gamma)]^2$$

Finally, the estimates $\hat{\gamma}_{cs}$ and $\hat{\gamma}_{ag}$ could be pooled by inversely weighting with regard to their estimated variances.[13]

For exact aggregation models, this kind of pooled estimation is discussed in detail by Jorgenson and Stoker (1986). In this case, with an individual regression model of the form

$$(3.43) \qquad E(y|x) = a(p,\gamma) + b(p,\gamma)'x \quad ,$$

the incorporation of cross section data into the estimation is particularly easy. Specifically, the estimation of (3.43) employs the cross section data through the OLS coefficients $\hat{a}$, $\hat{b}$ of y regressed on x and a constant, which consistently estimate $a(p_{t_0},\gamma)$ and $b(p_{t_0},\gamma)$. This represents a substantial computational simplification over (3.40) or (3.41) with a nonlinear individual model.

For later reference, it is useful to restate this feature of exact aggregation models in different terms. Because the aggregate equation from (3.43) is

$$(3.44) \qquad E_t(y) = \phi(p_t,E_t(x),\gamma) = a(p_t,\gamma) + b(p_t,\gamma)'E_t(x)$$

we have that the "aggregate effect" $\partial E_t(y)/\partial E_t(x) = \partial\phi/\partial E_t(x)$ at time t_0 is just $b(p_{t_0},\gamma)$, and that the cross section OLS slope vector $\hat{b}$ consistently measures that effect. This coincidence of cross section and aggregate coefficients is implied by the exact aggregation format, and could be statistically tested to check the specification of such a model

With a substantively nonlinear model and a fair sized cross section data base, the estimation indicated in (3.40) or (3.41) can involve extensive computation, making the overall estimation job considerably harder than just estimating parameters with aggregate data. We close this section by raising some connections that permit partial methods of pooling.[14] These methods are the nonlinear analogy to pooling based on coefficients with exact aggregation models.

To set this up, recall that the vector x can represent products and other transformations of the basic observed individual variables, and suppose that x is specified so that $E_t(x)$ parameterizes distribution movements. In particular, we can determine μ_t as $\mu_t = H(E_t(x))$ in the aggregate model (3.39), rewriting it as

$$E_t(y) = \int f(p_t, x, \gamma)\ \rho[x|H(E_t(x)]\ dx\ = \phi^*(p_t, E_t(x), \gamma). \tag{3.45}$$

The "aggregate effect" at time t_0 is $\partial E_t(y)/\partial E_t(x) = \partial\phi^*/\partial E_t(x)$ evaluated at time t_0.

The connection works as follows (Stoker (1986a); suppose that the "score" $\ell_i = \partial \ln\ \rho(x_i|\mu_t)/\partial\mu_t$ can be estimated for each x_i in the cross section at $t = t_0$. Suppose further that $\hat{d}$ are the slope coefficients of regressing y_i on x_i using ℓ_i as the instrumental variable:

$$\hat{d} = (\sum \ell_i x_i^T)^{-1}\ (\sum \ell_i y_i)\quad . \tag{3.46}$$

The result is that $\hat{d}$ consistently estimates the aggregate effect, as in

$$\text{plim}\ \hat{d} = \partial E_t(y)/\partial E_t(x). \tag{3.47}$$

This is true regardless of the form of the individual model.[15]

One could envisage situations where the cross section coefficients $\hat{d}$ could be used to extrapolate $E_t(y)$ in subsequent time periods, in a way that was robust to the specification of the individual model. When the model is fully specified, this result is useful for partial pooling in the estimation of γ, as $\hat{d}$ estimates $\partial\phi^*(p_{t_0}, E_{t_0}(x), \gamma)/\partial E_t(x)$. For instance, if γ_{cs}^* denoted the parameters determined if this effect were known, then estimation could be based on minimum distance, as in

$$(3.48) \qquad \hat{\gamma}_{cs}^{*} =$$

$$\text{argmin}_{\gamma_{cs}} [\hat{d} - \partial\phi^{*}(p_{t_0}, E_{t_0}(x), \gamma)/\partial E_t(x)]^T V_{\hat{d}}^{-1} [\hat{d} - \partial\phi^{*}(p_{t_0}, E_{t_0}(x), \gamma)/\partial E_t(x)]$$

where $V_{\hat{d}}$ is an estimate of the variance of $\hat{d}$. This objective function could also be included as part of a partial pooling procedure in the standard way. Moreover, while estimation of the scores ℓ_i may appear daunting at first glance, in leading cases they do not need estimating. For instance, if the distribution is in the exponential family form (3.24) with $D(x) = x$, then ℓ_i is proportional to x_i and $\hat{d}$ is the OLS regression coefficients of y_i on x_i. This form occurs if x is normally distributed, for instance. An example is given in our discrete choice example of (3.17-3.20), with Σ_t constant over time; there $\hat{d}$ is the OLS coefficient of the 0-1 variable y_i on $x_i = \ln M_i$, which consistently estimates the effect of changing the mean of log M on the proportion of purchasers, or $\partial E_t(y)/\partial \mu_t$ in our earlier notation. If distributional movements are represented as translations (or can be written so, as in proportional scaling), with $\rho_t(x) = \rho_0(x - E_t(x))$, then $\hat{d}$ can be based on nonparametric estimates of the scores. These types of estimates are known as "average derivative estimators", and are discussed in a different context in Stoker (1992).

4. Empirical Approaches that Account for Aggregation Over Individuals

Recent work has involved a wide variety of modeling approaches for studying the issues raised by aggregation over individuals. Our coverage of the theoretical considerations provides some central themes to discuss in each of these areas. We now turn to an area-by-area summary of different approaches.

4.1 Statistical Assessment of Distributional Effects in Macroeconomic Equations

Compositional effects must be present in aggregate data unless the marginal reactions of individuals are remarkably similar. We first consider work that looks in crude fashion to see where distributional effects are manifested in aggregate data. One way of making such comparisons is to contrast economic variables across situations where the distributional structures are grossly different. An older example of this kind of comparison is given by Franco Modigliani (1970), who explains differences in savings rates across countries by focusing on population growth rates and age, motivated by the notion that individuals in different countries will have similar needs for saving consistent with a simple life cycle.

More germane to standard macroeconomic analysis is the assessment of distributional effects over time in a particular economy. Simple approaches here amount to including distributional variables in a standard macroeconomic equation, and testing for whether they have a significant effect. An early example of this kind of study is by Alan Blinder (1975), who studied the effects of income distribution on average consumption in the U.S. Blinder included relative income distribution variables (quantiles), and failed to find any significant effects.

Blinder's pioneering results are of interest for several reasons, including pointing out two difficulties with measuring distributional effects. The first problem is that there may be too little variation in the distributional variables of interest over time, as with the relative income distribution in the United States. The second problem is that without some micro-macro correspondence in the modeling approach, even significant results may be difficult to interpret, aside from asserting that "distribution apparently matters." For instance, if Blinder had found significant effects of relative income quantiles, this would suggest consumption differences attached to the relative position of individual incomes, but not the income level, which would seem more relevant for individual consumption decisions. The interpretation issue is exacerbated for the inclusion of variables such as the Gini coefficient of the income distribution, which is not obviously traceable to individual income effects.

As indicated in Section 3.1, such difficulties of interpretation are addressed by including distributional statistics that are themselves averages, such as proportions of individuals in well-defined categories. The effects of such proportions are interpretable because they coincide exactly with dummy variable methods of studying individual differences. For instance, recall our earlier example of investigating small-large family differences in demand, and in particular, equations (2.3-5). With cross section data, one might take a first cut at looking at such differences by fitting the regression equation

$$y_{it} = a + b\, M_{it} + d\, A_{it} + u_{it} \qquad i = 1,\ldots,N_t \tag{4.1}$$

where $A_{it} = 1$ if family i is small, and $A_{it} = 0$ if large, and testing whether $d = 0$. The aggregate analog of this is to include the proportion of small families $P_{0t} = N_t^{-1} \sum A_{it}$ in the aggregate equation, as in

(4.2) $$\bar{y}_t = a + b \bar{M}_t + d P_{0t} + u_t$$

and testing for whether d = 0, where we have abstracted from price effects for simplicity. As in (2.5), it is clear that a measures the basic level of demands for large families, and d measures the difference between the levels for small and large families.

Recent efforts to characterize distributional effects using proportion variables have been more successful than their predecessors. Stoker (1986c) examines the robustness of the popular Stone-Geary linear expenditure system (LES) by including proportions of families in various ranges of the real income distribution as regressors. For discussing the results, consider a typical equation of this system (say for expenditure on commodity group 1), augmented for distributional effects, which takes the form

(4.3) $$\bar{y}_t = (1-b)\gamma_1 p_{1t} - \sum_{k\neq 1} b\gamma_k p_{kt} + b \bar{M}_t + a + \sum d_j P_{jt} + u_t$$

being linear in prices (p_{kt}) and average total expenditure $\bar{M}_t$, and where P_{jt} denotes proportions of families in fixed ranges of the real income distribution. From our discussion above, it is clear that the d_j coefficients pick up departures of the micro Engel curve from linearity, which coincide with distributional effects in the aggregate equation). Moreover, $d_j = 0$ for all j coincides with the linear expenditure system being statistically valid for each household as well as for the aggregate data.[16]

Three features of the empirical results of this study are of interest for our discussion. First, the hypothesis of no distributional effects ($d_j = 0$ for all j) is soundly rejected, and including the proportion variables substantially changed the estimates of marginal income effects. For example, for food, which is around a third of the budget, $\hat{b} = .1$ when the proportions were omitted, but $\hat{b} = .3$ when they were included. This gives evidence for heterogeneity in individual responses, as well as suggests that accounting for heterogeneity may bring macro parameter estimates more in line with estimates from micro data. Second, while distributional effects were clearly evidenced, the separate estimates of the d_j parameters were not precisely estimated. This coincides with the issue of little distributional variation, and forces the conclusion that detailed individual demand patterns are unlikely to be easily measured with aggregate data alone, even augmented by proportions. Full micro-macro modeling of the kind discussed in Section 3, and in sections below, appear necessary for a successful characterization of the impacts of individual heterogeneity in aggregate data.

The third feature of the results is the most intriguing, and suggestive of future research questions. In particular, a more conventional approach to assessing the LES would be to look for dynamic misspecification, and here, the original LES estimates displayed substantial serial correlation in the residuals. In fact, the estimation of a quasi-differenced formulation suggested that a first differenced (or cointegrated) LES model would be appropriate for the aggregate data, and the estimates of marginal income effects (b above) had intuitively reasonable values under this specification.

The intriguing feature arises from considering dynamic and heterogeneity influences simultaneously. In particular, no serial correlation was evidenced for the model with proportions. Neither the quasi-differenced model, nor the model in levels with proportions, were strongly rejected against a specification that permitted both heterogeneity and serial correlation. In other words, the model that accommodated individual heterogeneity in expenditure levels and a simple, first differenced dynamic model provided practically equivalent descriptions of the aggregate demand data. It is easy to see how this could happen, and the implications for aggregate data analysis are strong. Namely, suppose (4.3) provided a statistically decent representation of the individual heterogeneity, then first differencing it gives

$$(4.4) \qquad \bar{y}_t - \bar{y}_{t-1} = (1-b)\gamma_1 (p_{1t}-p_{1t-1}) - \sum_{k\neq 1} b\gamma_k (p_{kt}-p_{kt-1}) +$$

$$b (\bar{M}_t-\bar{M}_{t-1}) + \sum d_j (P_{jt}-P_{jt-1}) + u_t-u_{t-1} .$$

Since the income distribution evolves slowly, with the proportion differences $P_{jt}-P_{jt-1}$ negligible or nearly constant, differencing can effectively eliminate their impact. The broad point is that, because distributional effects naturally exist in aggregate data, distributional effects are primary candidates for the kinds of omitted features giving rise to aggregate dynamic structure. The interesting result is that accommodating individual heterogeneity may go some distance in explaining the source of apparent dynamics in aggregate data.

Stoker's study is flawed in a number of ways, such as the use of proportions of the real income distribution in place of proportions of the total expenditure distribution. More important, though, is that the use of the LES sets up a very restrictive "straw man" to shoot at. Exoneration of this system would be consistent with individual Engel curve patterns that are linear, which have never been observed in surveys of individual budgets.

In response to some of these concerns, Adolph Buse (1992) devises a similar testing strategy based on the Quadratic Expenditure System (QES) of Pollak and Terence Wales (1979), which permits quadratic micro Engel curves, and studies several kinds of dynamic specifications, such as those consistent with habit formation. Using Canadian data, Buse finds virtually the same results, which differ only to the extent that evidence is found for preferring the demand model with heterogeneity over dynamic demand specifications without heterogeneity. He concludes that the role of heterogeneity as well as its implications for dynamic structure were not due to the restrictive LES equations, nor are just an artifact of US demand data.

Individual households differ in many ways, and focusing on the income distribution may be a particularly ill-designed approach for studying distributional effects. Barring tumultuous times like civil revolutions, movements in income distributions tend to be quite smooth, which can preclude precise measurement of their impacts on aggregate variables. Distributions of other types of individual characteristics clearly exhibit more variation; the familiarity of "baby boom" and "bust" cycles to describe the U.S. post war experience raises the age distribution and family size distribution as natural candidates. In terms of the age distribution, this point is implemented by Ray Fair and Katherine Dominguez (1991). In particular, they find strong evidence of age distribution effects in four different kinds of traditional macroeconomic equations, including one for consumption. While they restrict the individual age impacts to have a quadratic shape, they are able to interpret the estimated age patterns in straightforward ways, via the (individual) age structure that they are associated with.[17]

While a useful starting point, the methods discussed above are admittedly crude, and implemented in an exploratory, or ad hoc, fashion. The broad message of this work is that applying crude, simple methods can find evidence of distributional effects in various settings, and permit comparisons with other estimation approaches. Distributional effects are not completely masked in the aggregate data studies discussed above, although the relative importance of individual heterogeneity versus common dynamic structure is an open question. At any rate, distributional variables are natural candidates for inclusion in tests of specification of any empirical macroeconomic equation. To get a closer assessment of the true individual structure, one needs to carry out more full micro-macro modeling, which dictates exactly how distributional influences and behavioral effects are to be separated. We now turn to work that has developed this paradigm in demand analysis.

4.2 Individual Heterogeneity and Distributional Effects in Demand Analysis

The majority of work done on modeling individual and aggregate data has been done in the context of studying demands for various commodities. In historical perspective, this work follows the introduction of flexible functional forms for representative agent demand models, which in turn follows the fairly widespread application of the (Stone-Geary) Linear Expenditure System to aggregate demands. Interest in demand models that accommodate individual heterogeneity is motivated by at least three basic features. First is the well-documented existence of demographic effects and nonlinearity of Engel curves in cross section data, or features that immediately imply the presence of distributional effects in aggregate demand. Second is the fact that until recently, the only source of information on reactions to varying prices was aggregate time series data. This meant that accounting for price, income and demographic effects required pooling of aggregate and individual data sources. Third, the application of demand systems to welfare analysis is extremely limited when based on aggregate analysis alone. Consumer surplus analysis, the standard aggregate method, is deficient in several ways. For instance, there are the well known theoretical issues of whether consumer surplus accurately measures equivalence or compensating variation, when a single family's demand has been measured. But more important is that differences in needs across families implies that welfare impacts will likewise differ, in ways that make the use of a single surplus measure at best ambiguous. The only consistent way of constructing a single welfare measure is to implement an explicit social welfare function, but this requires realistic individual demands and/or preferences as inputs.

A logical starting point for our discussion is with demands that are linear in the total budget

$$(4.5) \qquad y_{it} = p_{1t}q_{1it} = a(p_t) + b(p_t)M_{it}$$

where p_{1t}, q_{1it} are the price and quantity of a good (say # 1). Preferences that give rise to demands of this form are characterized by Gorman (1961), and include the Linear Expenditure System and similar models; see Blackorby, Richard Boyce and Russell (1978), among others. Such linear structures, with common marginal reactions, have been used in other modeling contexts as well; for instance, see the consumption model of Martin Eichenbaum, Lars Peter Hansen and Scott Richard (1987).[18]

The first direct use of distributional information in aggregate demands arises from incorporating nonlinearity in income effects.[19] The principal examples arise from models where budget shares vary with log total expenditures $\ln M_{it}$.[20] Aggregate budget shares in these models depend on the entropy statistic

$$(4.6) \qquad \mathcal{E}_t = \frac{\sum M_{it} \ln M_{it}}{\sum M_{it}} \qquad t = 1,\dots,T.$$

Ernst Berndt, Masako Darrough and Diewert (1977) implement a version of translog demand equations (discussed below) of this form. Another popular demand model in this form is Deaton and Muellbauer's (1980a,b) "Almost Ideal" or AIDS demand system. Each equation from this system takes the form

$$(4.7) \qquad w_{1it} = \alpha_1 + \sum_j \gamma_{1j} \ln p_{jt} - \beta_1[\ln C(p_t)] + \beta_1 \ln M_{it} + \epsilon_{1it}$$

where

(4.8) $$C(p_t) = \exp[\alpha_0 + \sum_j \alpha_j \ln p_{jt} + (1/2) \sum_j \sum_k \gamma_{jk} \ln p_{jt} \ln p_{kt}].$$

and ϵ_{1it} is an additive disturbance. The associated market budget share equation is

(4.9) $$W_{1t} = \alpha_1 + \sum_j \gamma_{1j} \ln p_{jt} - \beta_1(\ln C(p_t)) + \beta_1 \mathcal{E}_t + \epsilon_{1t}$$

where $\epsilon_{1t} = \sum M_{it}\epsilon_{1it}/\sum M_{it}$ is the aggregate disturbance. The parameters of this model are restricted by integrability conditions; see Deaton and Muellbauer (1980a,b) for details. For estimation, the complicated nonlinearity in parameters is often sidestepped by replacing $C(p_t)$ by an observed price index.

Proper implementation of this model involves observing the statistic $\mathcal{E}_t$ for each time period. The early applications discussed above sometimes used a distribution restriction so that $\ln \bar{M}_t$ can be used in place of $\mathcal{E}_t$. In particular, we have that

(4.10) $$\mathcal{E}_t = \ln \bar{M}_t + \tilde{\mathcal{E}}_t$$

where $\tilde{\mathcal{E}}_t$ is Theil's entropy measure of relative income inequality

(4.11) $$\tilde{\mathcal{E}}_t = \frac{\sum M_{it} \ln (M_{it}/\bar{M}_t)}{\sum M_{it}}$$

Under the distributional assumption that $\tilde{\mathcal{E}}_t = \tilde{\mathcal{E}}$ is constant over t, then the aggregate model takes the form

(4.12) $$W_{1t} = \tilde{\alpha}_1 + \sum_j \gamma_{1j} \ln p_{jt} - \beta_1(\ln C(p_t)) + \beta_1 \ln \bar{M}_t$$

where $\tilde{\alpha}_1 = \alpha_1 + \beta_1\tilde{\mathcal{E}}$. This assumption is used in Deaton and Muellbauer's (1980a) estimation, and is consistent with "proportional scaling," where all individual expenditure values M_{it} just scale up or down proportionately with mean total expenditure $\bar{M}_t$ over time.[21]

The obvious similarity between (4.12) and (4.7) can be mistaken as justifying a per capita, representative agent model for demands. Equation (4.12) arises from a definite individual level model and employs a distribution restriction for aggregation, which *coincidentally* gives the same estimation equation as a AIDS model applied for a representative agent. In particular, (4.12) rests on the assumption that a) (4.7) is valid, with no individual heterogeneity in demands aside from income effects and b) that relative entropy $\mathcal{E}_t$ is constant over time. Each of these assumptions is testable with micro data, and patently unrealistic; but for our purposes we note that the parameter interpretations and integrability restrictions applicable to (4.12) come directly from (4.7). This notion of what aggregation structures give rise to equations analogous to those fit in a "representative agent" approach has been studied by Lewbel (1989b).[22]

The joint distribution of total expenditure and family demographic variables is incorporated in the translog model of Jorgenson, Lau and Stoker (1982). A budget share equation from this system takes the form

$$(4.13)) \quad w_{1it} = \left(\frac{1}{D(p_t)}\right) (\alpha_1 + \sum_j \beta_{1j} \ln p_{jt} + \beta_1 \ln M_{it} + \sum_s \beta_{As} A_{sit}) + \epsilon_{1it}$$

where A_{sit}, $s = 1,\ldots,S$ are 0-1 variables indicating demographic structure of the family, and $D(p_t) = -1 + \sum_k \sum_j \beta_{kj} \ln p_{jt}$. As before, integrability restrictions are applicable to the parameters of this model. The associated market budget share is

$$(4.14) \quad W_{1t} = \left(\frac{1}{D(p_t)}\right) (\alpha_1 + \sum_j \beta_{1j} \ln p_{jt} + \beta_1 \mathcal{E}_t + \sum_s \beta_{As} PM_{st}) + \epsilon_{1t}$$

where $\mathcal{E}_t$ is the entropy term above, and

$$(4.15) \qquad PM_{st} = \frac{\sum M_{it} A_{sit}}{\sum M_{it}} ,$$

or, in words, PM_{st} is the proportion of total expenditure accounted for by families with $A_{sit} = 1$. Therefore, the market demand equation (4.14) has a size distribution effect $\mathcal{E}_t$, and demographic heterogeneity effects through PM_{st}.

Jorgenson, Lau and Stoker (1982) implement model (4.14) using observations over time on the distributional statistics $\mathcal{E}_t$ and PM_{st} for five demographic categories (family size, age of head, region and type of residence, and race of head), using 18 dummy variables A_{sit}. In principle, all parameters of the model could be estimated with aggregate data (including the distributional statistics) alone, but modeling a substantive number of demographic influences necessitates pooling aggregate data with other data sources. They use cross section data to estimate model (4.13) (for given value of price p), or to estimate the income and demographic effects, and pool those results with estimates of model (4.14) from aggregate data. This amounts to estimating price effects with data on varying prices over time, and income and demographic effects with data across individuals. The basic model indicates how estimates from different types of data sources are to be consistently combined.[23]

Instead of using data on the distributional statistics, it is clear that distributional restrictions could have been applied to generate a simple aggregate equation; for instance, i) $\mathcal{E}_t$ constant over time and iia) PM_{st} constant over time, gives market budget shares depending on only p_t and $\ln \bar{M}_t$, whereas i) and iib) $PM_{st}/\bar{A}_{st}$ constant over time, gives an aggregate equation of the form (4.13) with M_{it} replaced by $\bar{M}_t$ and A_{sit} replaced by $\bar{A}_{st}$. As discussed before, $\mathcal{E}_t$ and the relative proportions $PM_{st}/\bar{A}_{st}$ are "aggregation factors," whose constancy give the relevant distributional restrictions for motivating simple forms of aggregate equations.

While the inclusion of demographic characteristics gives a substantial generalization over models that omit them, evidence from individual cross section data shows that income and demographic effects are more complicated than those depicted translog equation (4.13). For instance, Lewbel (1991) and Jerry Hausman, Whitney Newey and Powell (1992) find evidence of more elaborate income structure than just one log expenditure term. Martin Browning (1992) surveys work that shows substantial interactions between income structure, family size and other demographic effects.

The ideal empirical situation for studying income, demographic and price structures of individual household demand would be based on an extensive panel survey, covering demand purchase across a large number of families and a large number of time periods. Coming close to this ideal is the recent study of Blundell, Pashardes and Weber (1992), who analyze the repeated annual cross section data bases from 1970-1984 of the British Family Expenditure Survey, involving 61,000 observations on household demands. For a seven good demand model, they find a quadratic version of the AIDS model (with $(\ln M)^2$ terms) to be adequate, including extensive coverage of individual demographic attributes. They also use the notion of constant "aggregation factors" as discussed above to develop a cohesive empirical explanation of how aggregate demand, aggregate total expenditure and price patterns can adhere to a fairly simple model over 1970-1984. In essence, they conclude that heterogeneous demographic influences are paramount and the income structure of the original exact aggregation models require some generalization. Moreover, with a proper accounting for distributional effects, parameter estimates correspond to those from micro data studies, and the aggregate demand model more accurately tracks aggregate demand data patterns simpler per-capita (representative agent) demand equations. This study sets the current standard for careful empirical work on the impact of aggregation in the study of demand behavior; as extensive data bases of this kind become available in other fields, similar studies would likewise be quite valuable.

While models that consistently treat individual household and aggregate demand behavior involve more extensive modeling than simpler, representative agent approaches, they are likewise more informative in applications, such as assessing alternative policy scenarios. Models of this kind can be used to forecast demands by different kinds of households, and assess the differential welfare impacts across different kinds of households. Stoker(1986b) uses the translog model (4.13, 4.14) in a retrospective analysis of the welfare impacts of the energy price changes of the 1970's, along similar lines to the early application of Jorgenson, Lau and Stoker (1980). This kind of application can be taken one step further, by combining individual welfare impacts via an explicit social welfare function, to get overall "good" or "bad" assessments. Jorgenson and Slesnick (1984) formulate an explicit social welfare function, and assess the implications of various policies on the pricing of natural gas using explicit interpersonal comparisons. While any specific method of combining individual welfare measures is subject to debate, it is clear that a full accounting of individual differences is necessary to get a realistic depiction of the benefits and costs of economic policy.

Work on demand analysis represents the most extensive development of models that account for aggregation over individuals, in terms of theoretical consistency and empirical properties. While exact aggregation models have appeared in other applied areas, recent empirical work has often approached the problems of aggregation from different directions, and used somewhat different methods. Moreover, as discussed by Blundell (1988), much current (micro level) demand modeling deals with situations where intrinsically nonlinear models are necessary (such as rationing and corner optima). We now discuss some of these other approaches and methods.

4.3 Aggregation and Goodness-of-Fit Tests

As discussed above, it is entirely possible for there to exist substantial heterogeneity in individual responses together with a simple, possibly linear relationship existing between the associated aggregates. While this setting immediately raises doubts as to the interpretation of the aggregate relationship, one could ask whether the aggregate equation could serve as a good tool for prediction. This amounts to renouncing any possible behavioral interpretation of such an equation, and justifying such aggregate equations through the need for parsimony in a larger modeling context. This kind of approach was laid out for linear models by Yehuda Grunfeld and Zvi Griliches (1963), who also give an early portrayal of distribution restrictions as a "synchronization" of individual responses. A revival and extension of these ideas is contained in Hashem Pesaran, Richard Pierce and M.S. Kumar (1989), who develop such a "goodness-of-fit" test in modern econometric terms.[24]

This idea can be seen easily as follows. Suppose that the population consists of N individuals (or groups), with the behavior of individual i given by the linear model

$$y_{it} = \alpha_i + x_{it}'\beta_i + u_{it} \qquad i = 1,\dots,N \ , \ t = 1,\dots,T \ . \tag{4.16}$$

x_{it} represents the principal economic variables of interest, with all individual differences captured by the coefficients α_i, β_i (any similarities or dissimilarities across individuals are left unspecified). This model implies the aggregate equation

(4.17) $$\bar{y}_t = N^{-1} \sum \alpha_i + N^{-1} \sum x_{it}'\beta_i + \bar{u}_t$$

This model can be implemented by estimating α_i and β_i for each individual, and inserting the estimates in the aggregate model (4.17); or adding up the individual equations for aggregate prediction.

The question of interest here is whether a simple model among aggregates could be estimated, namely

(4.18) $$\bar{y}_t \cong \tilde{\alpha} + \bar{x}_t'\tilde{\beta} + \tilde{u}_t \qquad t = 1,\ldots,T.$$

that would give the same degree of fit to the aggregates as the true model (4.17). A test of this situation (termed "perfect aggregation" by Pesaran, Pierce and Kumar (1989)) is a test of the restrictions

(4.19) $$N_t^{-1} \sum \alpha_i + N_t^{-1} \sum x_{it}'\beta_i = \tilde{\alpha} + \bar{x}_t'\tilde{\beta}, \quad t = 1,\ldots,T$$

performed with panel data (y_{it}, x_{it} for all i, t), using estimates of each of the coefficient values. Failure to reject this condition "justifies" the use of equation (4.18) in terms of aggregate goodness-of-fit.

The evaluation of this approach involves assessing the appropriateness of the linear micro models (4.16), as well as the results of the "goodness-of-fit" test. For the latter, consider the situation where such a test fails to reject, with (4.18) giving a statistically adequate depiction of the aggregate data patterns relative to the true micro model (4.16). What does this say? Consider the notion of 'aggregation factors' here; namely write the true model (4.17) as

(4.20) $$\bar{y}_t = \bar{\alpha} + \bar{x}_t'b_t + \bar{u}_t$$

where

(4.21) $$b_t = [\sum x_{it}\beta_i / \sum x_{it}] \quad .$$

With sufficient variation in $\bar{x}_t$ over time, (4.18) amounts to having b_t constant, or $b_t = \bar{\beta}$ for each time period. This clearly occurs in the exact aggregation case of constant coefficients, where $\beta_i = \beta_j$ $(= \bar{\beta})$ for each i, j. But in other cases, there are practical questions arising from the fact that b_t is based on the unobserved micro parameters and the distribution of x_{it} in each time period, and knowing that b_t is constant does not reveal what aspects of the distributional underpinnings are important.

For example, suppose that the estimation of the individual coefficients of (4.16) revealed that a group of micro agents had "large β's" and another group had "small β's". If b_t is constant for all t, then one can only conclude that the large-small differences are sufficiently smeared in the aggregate data as not to be noticed empirically. This is an unfortunate "synchronization" of x_{it} and β_i, as one cannot learn whether the data has involved a sectoral trend from "small" to "large" groups or vice versa, which is necessary information for applying the model out of sample. This issue has a simple answer, which is to model differences among β_i's using observable micro data, so that the aggregate model reflects as many systematic features of individual behavior as possible. Modeling all coefficient differences in this way amounts to an exact aggregation approach, with the "aggregation factors" based on observable features only.

Another depiction of the "synchronization" phenomena is given in Clive Granger's (1987,1990) analysis of "common factors." This work points out how studies of individual level data involve different sources of variation from studies of aggregate level data, as follows. Consider an individual level model of the form

$$(4.22) \qquad y_{it} = \alpha + \gamma p_t + \beta_0 x_{it} + \beta_1 x_{it}^2 + \epsilon_{it}$$

where p_t is a common observed variable as before, x_{it} varies over individuals and over time, and ϵ_{it} is a disturbance, uncorrelated over individuals and time. Suppose for simplicity that the variance of x_{it} in the population at time t is Σ^2, constant over time t. The aggregate $E_t(y)$ in a large population is

$$(4.23) \qquad E_t(y) = \alpha + \beta_1 \Sigma^2 + \gamma p_t + \beta_0 E_t(x) + \beta_1 E_t(x)^2$$

Rewrite the individual level model at a specific time, say t = 0, as

$$(4.24) \qquad y_{i0} = (\alpha + \beta_1 \Sigma^2 + \gamma p_0 + \beta_0 E_0(x) + \beta_1 E_0(x)^2)$$

$$+ \beta_0 [x_{it} - E_0(x)] + \beta_1 [x_{it}^2 - E_0(x)^2 - \Sigma^2] + \epsilon_{it}$$

Now, defining p_t, $E_t(x)$ (and $E_t(x)^2$) as "common factors," they are seen as the source of variation of the aggregate $E_t(y)$ over time t. Alternatively, the cross section variation of y_{it} at time t = 0 is due entirely to the deviation terms involving x_{it} and x_{it}^2 above. As such, the sources of variation are orthogonal in a natural way. For the aggregate model, the relevant "synchronization" of x_{it} values is through the common factors appropriate for the model.

This example underscores the idea of pooling individual level data and aggregate data; clearly both sources of variation apply to estimation of β_0 and β_1, and more precise estimation of these parameters can lead to more precise estimation of α and γ from the aggregate model. Granger (1987,1990) also argues that aggregate relationships can become more "linear," however, this argument does not appear applicable above, and therefore would need to be addressed in specific examples.[25]

The essential point here is that much is missed by focusing on aggregate equations alone, whether oversimplified or not. Aggregate "goodness of fit" tests of the kind outlined above can and should be performed as part of checking all restrictions of a micro-macro model, but not the only part. If there are economic reasons for individual behavioral differences that are not adequately captured in the micro model, then the aggregate level model suffers from important omissions, regardless of how well it fits aggregate data patterns. When individual differences are incorporated, estimation can involve entirely different sources of variation from individual level data and aggregate level data, however, the basic model dictates how those sources of variation can be combined. A proper justification for an aggregate model requires ruling out the omission of important individual differences, and the aggregate data alone may have little to say about this.

4.4 Time Series Analysis and Dynamic Behavioral Models

We have stressed above how the interplay between individual heterogeneity and dynamic structure raises many basic issues for the modeling of aggregate data. The well established empirical tradition of measuring short run and long run effects, as well as judging transitory and permanent impacts for forecasting, underscore the practical importance of assessing the impact of individual heterogeneity in dynamic equations estimated with aggregate data. There has been relatively little attention to these issues, with some notable exceptions (see Granger (1990)). We now discuss some of the issues, to place them in the context of our survey.

We begin by considering difficulties in interpreting dynamic equations estimated with aggregate data. The issue here is that aggregation over heterogeneous (individual) time series processes tends to result in longer, more extensive processes applicable to aggregate data. This general notion is in line with the ideas of heterogeneity giving rise to observed dynamics discussed in Section 4.1, and a general discussion of heterogeneity and lag structure is given in Pravin Trivedi (1985). Here we illustrate these ideas using a simple example of the form recently studied in Marco Lippi (1988) and Lewbel (1992).

Suppose that we are studying an economy of N individuals, and that the model applicable to individual i is an AR(1) process of the following form

$$y_{it} = \alpha + \gamma_i y_{it-1} + \beta z_{it} + \epsilon_{it} \tag{4.25}$$

where z_{it} is a set of predictor variables, and the first order coefficient γ_i varies over individuals. The aggregate model in a large population is therefore

$$\bar{y}_t = \alpha + N^{-1} \sum \gamma_i y_{it-1} + \beta \bar{z}_t \quad . \tag{4.26}$$

Because equation (4.25) applies for y_{it-1}, it is impossible to treat γ_i and y_{it-1} as uncorrelated (unless $\gamma_i = \gamma$ for all i). In particular, by recursive substitution of (4.25) into the expression for $N \sum \gamma_i y_{it-1}$, the aggregate model (4.26) is rewritten as

$$\bar{y}_t = \alpha + \Gamma_1 \bar{y}_{t-1} + \Gamma_2 \bar{y}_{t-2} + \Gamma_3 \bar{y}_{t-3} + \dots + \beta \bar{z}_t \tag{4.27}$$

where the aggregate lag coefficients are

$$\begin{aligned} \Gamma_1 &= E(\gamma_i) \\ \Gamma_2 &= E(\gamma_i^2) - E(\gamma_i)^2 = Var(\gamma_i) \\ \Gamma_3 &= E(\gamma_i^3) - 2E(\gamma_i)E(\gamma_i^2) + E(\gamma_i)^3 \end{aligned} \tag{4.28}$$

and so forth, where Γ_j, $j \geq 3$, is determined by the first through j^{th} moments of the distribution of γ_i in the population. Therefore, under the individual model (4.25), the low order moments of the distribution of first-order coefficients γ_i can be solved for from estimates of the Γ_j parameters. Setting aside the natural modeling questions of whether the Γ_j coefficients have stable structure for large j, or what lag length is appropriate in practical applications, this example illustrates how individual differences can generate more complicated aggregate dynamics. Obviously, the same (lag-lengthening) phenomena would occur if (4.25) displayed a more complicated lag structure than AR(1).

For a bit more clarity on the differences between the individual and aggregate level models, imagine one is studying consumption expenditures C_{it}, and that the economy consists of two kinds of households. The first household type ($A_i = 0$) is headed by irresponsible yuppies who spend every cent of current earnings (I_{it}), following the model

$$C_{it} = I_{it} \qquad A_i = 0 \tag{4.29}$$

The second household type ($A_i = 1$) is headed by uninteresting stalwarts who formulated a life plan while in high school, took jobs with secure earnings, and implemented perfect consumption smoothing. Setting aside real interest rate effects, these households follow the model

$$C_{it} = C_{it-1} \qquad A_i = 1 \tag{4.30}$$

These models are combined into an exact aggregation model as

$$C_{it} = A_i C_{it-1} + (1-A_i) I_{it} \quad , \tag{4.31}$$

and the correct aggregate model takes the form[26]

$$(4.32) \qquad \bar{C}_t = N^{-1} \sum A_i C_{it-1} + N^{-1} \sum (1-A_i) I_{it} \quad .$$

Obviously mean current consumption depends on the distributional structure of the population, namely the lagged consumption of stalwarts and the earnings of yuppies.

However, suppose that this current average consumption were studied as a function of average earnings and lagged consumption values. Equation (4.32) is in the form (4.25) with $y_{it} = C_{it}$, $\alpha = 0$, $\gamma_i = A_i$, $\beta = 1$ and $z_{it} = (1-A_i)I_{it}$. Supposing that the population is evenly split between stalwarts and yuppies, and that mean earnings at time t is the same for each group, the aggregate equation (4.27) takes the form

$$(4.33) \qquad \bar{C}_t = .5\, \bar{C}_{t-1} + .25\, \bar{C}_{t-2} + .125\, \bar{C}_{t-3} + \ldots + .5\, \bar{I}_t$$

The point is that the dynamics evidenced in this equation are nothing like the dynamics exhibited by either stalwarts or yuppies. On the basis of aggregate data alone, one could not distinguish our artificial setup from one with a common individual model of the form

$$(4.34) \qquad C_{it} = .5\, C_{it-1} + .25\, C_{it-2} + .125\, C_{it-3} + \ldots + .5\, I_{it}$$

which exhibits fairly slow adjustment for every household. Moreover, if the composition of stalwarts and yuppies were time varying, then the coefficients of (4.33) would likewise be time varying. Of course, the basic problem lies in trying to give a behavioral interpretation to the dynamic equation (4.33). The proper model is (4.32), which would reveal the stalwart-yuppie heterogeneity from the effects of the right-hand variables, by capturing the compositional effects in a way consistent with the correct individual model.

These kinds of interpretation issues may be particularly pronounced in studies of durable goods. For instance, suppose that the aggregate stock of refrigerators grew quite gradually over time, then it is natural to expect that an aggregate equation with several lags would describe the evolution of this stock. But at the individual level, adjustment occurs differently: people buy a new refrigerator a discrete times; when their old one breaks, or when they change decor as part of moving to new house, etc. In any case, it is problematic to ascribe much of a behavioral interpretation to a time series model describing the aggregate stock of refrigerators.

It is one thing to point up difficulties in casual behavioral interpretations of equations estimated with aggregate data, but it is quite another to make constructive remarks on aggregation relative to dynamic behavioral models, such as models of individual choice under uncertainty. While the literature is replete with applications of such models directly to aggregate data (under the assumption of a representative agent), we can ask what issues arise for modeling aggregates if such a behavioral model is applied to individual agents themselves. Since models that account for uncertainty involve planning for the future, a realistic consideration of heterogeneity must include all differences relevant to planning; namely differences in objectives (tastes, technology, etc.) as well as differences in the information used in the individual planning processes. Another central issue concerns the implications of markets that can shift uncertainty across agents, such as insurance or futures markets.

It is fair to say that the development of macroeconomics over the last two decades has been preoccupied with issues of uncertainty, and we cannot do more than just touch the surface of these issues here. Nevertheless, it is informative to look at a familiar paradigm from our vantage point. For this we consider intertemporal consumption models as popularized by Robert Hall (1978) and Hansen and Kenneth Singleton (1982).

We spell out the general setting first, and then give specializations.

Family i chooses how much to spend C_{it} at time t as part of a plan that takes account of future uncertainty in wage income and other features, by optimizing over time with regard to available wealth. Specifically, at time t family i's consumption plan arises from maximizing expected utility

$$(4.35) \qquad E_{it}[\sum_{\tau=0}^{T_i} \delta_i^{\tau} U_{it+\tau}(C_{it+\tau}) | \mathcal{I}_{it}]$$

subject to the amount of available wealth, where future wages and other income are not known with certainty. Here $U_{it+\tau}()$ is the utility of consumption for family i's at time $t+\tau$, δ_i its (subjective) discount factor and T_i sets the planning horizon. E_{it} reflects family i's expectation at time t, where expectations are formed with the information available at time t, denoted as $\mathcal{I}_{it}$. Consider the planning over periods t and t+1, where one could earn (possibly uncertain) interest κ. Optimal planning will equate (properly discounted) marginal utilities, giving the (Euler) equation $E_{it}([\delta_i/(1+\kappa)] U_{it+1}'(C_{it+1})|\mathcal{I}_{it}] = U_i'(C_{it})$, which we rewrite as

$$(4.36) \qquad [\delta_i/(1+\kappa)]U_i'(C_{it+1}) = U_i'(C_{it}) + V_{it+1}$$

where $V_{it+1} = [\delta_i/(1+\kappa)] U_i'(C_{it+1}) - E_{it}[R_1 U_i'(C_{it+1})|\mathcal{I}_{it}]$. The theory states that any departures of spending from the plan must be unanticipated, namely that $E_{it}(V_{it+1}|\mathcal{I}_{it}) = 0$. In other words, V_{it+1} reflects adjustment in reaction to "news" not known at time t. One may assess their job is more secure because of a surprise upswing in the economy, Uncle Ned could hit the lottery, or one could learn that a youngster in the family has a serious, costly illness. The behavioral theory just states that each family plans the best they can, and adjusts as new events unfold.

For some immediate implications, we begin with Hall's simplification of this model. Suppose that the interest rate κ is known; $\kappa = r$, and that family preferences are identical and quadratic; $U_{it}(C_{it}) = - 1/2 (B - C_{it})^2$, with B a "bliss" value of spending, and $\delta_i = \delta$. Solving (4.36) for C_{it+1} gives

$$(4.37) \qquad C_{it+1} = [1-\delta^{-1}(1+r)]\, B + \delta^{-1}(1+r)\, C_{it} + v_{it+1}$$

$$= a + b\, C_{it} + v_{it+1}$$

where $a = [1-\delta^{-1}(1+r)]\, B$, $b = \delta^{-1}(1+r)$ and $v_{it+1} = -\delta^{-1}(1+r)\, V_{it+1}$. Here, spending by family i at time t+1 is a linear function of spending at time t, plus any adjustment due to unanticipated events; with $E_{it}(v_{it+1}|\mathcal{I}_{it}) = 0$.[27] If each family discounts utility at the rate of interest; $\delta^{-1}(1+r) = 1$; then spending follows the familiar "random walk" $C_{it+1} = C_{it} + v_{it+1}$.

If the economy consists of N families, then average consumption is

$$(4.38) \qquad \bar{C}_{t+1} = a + b\, \bar{C}_t + \bar{v}_{t+1} \quad .$$

Suppose that $\mathcal{I}_t$ denotes information that every family has at time t (we assume there is some), then the planning theory asserts that $E_t(\bar{v}_{t+1}|\mathcal{I}_t) = N^{-1} \sum E_{it}(v_{it+1}|\mathcal{I}_t) = 0$. At this stage, differences in $\mathcal{I}_{it}$ across families, (heterogeneity in information) has little effect, only limiting the stochastic restrictions implied on the aggregate adjustment. Recoverability applies here: estimates of δ and B can be derived from estimates of $a = [1-\delta^{-1}(1+r)]B$, $b = \delta^{-1}(1+r)$.

Heterogeneity in preferences can be modeled in the same fashion as with our earlier discussion. For example, with quadratic preferences, suppose that times when greater spending are required are adequately modeled by raising the bliss point B in preferences; specifically $U_{it}(C_{it}) = -1/2\, (B_{it} - C_{it})^2$, where larger B_{it} indicates higher need for expenditure by family i at time t. Finally, for notational simplicity, suppose that at any time families are either needy ($A_{it} = 1$) or not ($A_{it} = 0$), with the bliss point modeled as $B_{it} =$

$\alpha + \gamma A_{it}$. Solving (4.36) now gives spending by family i as

$$(4.39) \quad C_{it+1} = B_{it+1} - \delta^{-1}(1+r) B_{it} + \delta^{-1}(1+r) C_{it} + v_{it+1}$$

$$= [1-\delta^{-1}(1+r)] \alpha + \gamma A_{it+1} - \gamma\delta^{-1}(1+r) A_{it} + \delta^{-1}(1+r) C_{it} + v_{it+1}$$

$$= a + c A_{it+1} + d A_{it} + b C_{it} + v_{it+1}$$

Average consumption is

$$(4.40) \quad \bar{C}_{t+1} = a + c P_{t+1} + d P_t + b \bar{C}_t + \bar{v}_{t+1} ,$$

where P_t denotes the proportion of families with higher needs ($A_{it} = 1$) at time t. The same stochastic restrictions apply to $\bar{v}_t$ as before, and the basic model parameters α, δ and γ are (over) identified by a, b, c and d.

We have used the simple "needy or not" distinction for illustration, as it is easy to see how this model could be derived for a more detailed scheme of planning for various things; feeding and clothing teenagers, college spending, or reduced spending in retirement; especially given their obvious connection with observable demographic attributes (age, family size, etc.). The resulting model would express average current spending $\bar{C}_{t+1}$ in terms of past spending $\bar{C}_t$, and the demographic structure of the distribution of families, as relevant to the lifetime spending plan. As above, such a model would be applicable to data on individual families as well as aggregate data.

While we have shown how individual heterogeneity can be accounted for in studying intertemporal consumption, we round out our discussion with two further observations. First, intrinsic nonlinearities cause complications for aggregation here, as in other areas. Suppose that interest rates are known and families have identical preferences; $U_{it}() = U()$, but that marginal utility U' is an invertible, nonlinear function of spending C. Following our above logic gives aggregate consumption as

(4.41) $$\bar{C}_{t+1} = N^{-1} \sum U'^{-1}[(1+r)\delta^{-1} U'(C_{it}) + v_{it+1}] \ .$$

The nonlinearity of U' requires $\bar{C}_{t+1}$ to depend explicitly on the distribution of C_{it} and v_{it+1} across all families. The behavioral theory asserts only that v_{it+1} is unanticipated, and thus uncorrelated with C_{it} in family i's forward planning process. Much more distributional structure is necessary for an adequate specification of this kind of model to be used for analyzing average consumption data. Heterogeneity in preferences complicates this further.

One source of such additional structure is appealed to in many macroeconomic studies, namely the existence of complete efficient markets. For instance, if families are further assumed to have identical homothetic preferences, Mark Rubinstein (1974) has shown how the presence of efficient markets implies that all idiosyncratic risk will be optimally shared, with family i's consumption a stable multiple of average consumption: $C_{it} = \theta_i \bar{C}_t$ for all i, where $N^{-1} \sum \theta_i = 1$. Homotheticity of preferences implies that marginal utility factors as $U'(\theta C) = a(\theta)\ U'(C))$, so that the Euler equation for family i

(4.42) $$E_{it}\{[\delta/(1+r)]\ U'(C_{it+1}) | \mathcal{I}_t] = U'(C_{it})$$

holds for average consumption, namely

(4.43) $$E_t\{[\delta/(1+r)]\ U'(\bar{C}_{t+1}) | \mathcal{I}_t] = U'(\bar{C}_t),$$

since $U'(C_{it}) = U'(\theta_i \bar{C}_t) = a(\theta_i)\ U'(\bar{C}_t)$ for all i and t. What is going on here is that optimal risk sharing implies that the individual equation (4.42) is a proportional ($a(\theta_i)$) copy of the same equation for the average consumption. In other words, the efficiency of insurance, futures and other markets acts to removes the impact of individual heterogeneity in information and risk. With identical homothetic preferences, each family plans expenditures in line with average consumption.

The argument that markets are sufficiently efficient to erase concerns about individual differences has been used elsewhere; for instance Gary Hansen 1985 raises the notion that a worker becoming unemployed could result from a process akin to a random lottery; prior to the lottery, the planning by all individuals is identical. Whether markets and/or institutions of this level of efficiency actually or approximately exist is debatable, and we will not discuss the available scientific evidence.

But differences in the needs and plans of individual families are evident, and in this context, it is important to stress how the individual behavioral model is logically distinct from coordination invoked by market interactions. Under the assumptions giving equation (4.39), equation (4.40) holds. Coordination across families induced by market interactions may permit (4.40) to be simplified, or may not. Consequently, building a realistic model does not involve a choice between accounting for individual heterogeneity *or* efficient markets; individual heterogeneity in behavior must be accounted for first, and the role of market interactions assessed subsequently.[28]

4.5 Market Participation and Other Models of Discrete Responses

While markets for insurance can serve to lessen heterogeneity in individual planning processes, there are many other roles that markets can play in aggregate data. Market participation models focus on a more primitive role, which is to account explicitly for the fact that individual households are choosing whether to buy a product, and that individual firms are choosing whether to produce the product. The "in or out" decision is binary in character, and determined by the prevailing level of market prices. In this setting, the price level determines what fraction of the consuming or producing population is active in the market, as the aggregate impact of heterogeneous, extensive margin decisions. This contrasts with our treatment of prices in continuous spending decisions, where they enter as common variables for all households.[29]

We can again appeal to our discrete choice example of Section 3.2 for illustration. In particular, the individual model (3.17) states whether a household purchases at price p_t, or participates in the market, and the aggregate equation (3.20) specifies what fraction of the population is participating. While we discussed this model in terms of aggregation over the income distribution, it is equivalently cast as a model of choice at various price levels. The overall issue is familiar to students of the microeconometric literature, as any treatment of selection bias has the same structure.[30]

A very natural setting for this kind of model is the study of employment. Here the decisions of whether to participate (getting a job) are made by potential workers comparing offered wages to reservation wages (or in current times of business restructuring, participation may be determined by firms offering positions). Thomas MaCurdy (1987) spells out how to build this kind of model of labor supply. In his setup, the employment participation percentage is modeled via an aggregated probit discrete response model.

A full implementation of an aggregate participation model is given in Heckman and Guilherme Sedlacek's (1985) estimation of a two sector model of labor markets. Here selection occurs between two labor markets, with the analysis permitting estimation of the wage effects of various individual skills, and employs lognormal distribution assumptions on unobserved wage differences. While this model treats capital across the sectors somewhat casually, this study is notable in that the authors give a convincing verification of the basic distributional assumptions used.

Another kind of aggregate participation model used recently is the short-run industry production model originally proposed by Hendrik Houthakker (1955). This kind of model involves aggregation over fixed input-output technologies, where participation is determined by whether profits are nonnegative at prevailing input and output price levels. The Houthakker setup is as follows: suppose individual production facility i can produce one unit, using a_{1i} units of labor, and a_{2i} units of capital, where the production requirements (a_{1i},a_{2i}) vary over the potential producers of the product. Production unit i will produce if its short term profits are nonnegative, or $p - wa_{1i} - ra_{2i} \geq 0$, where p is the price of the output, and w, r the input prices. Let $\mathcal{A}(w/p,r/p) = \{i \mid 1 - (w/p)\, a_{1i} - (r/p)\, a_{2i} \geq 0\}$ denote the set of units with nonnegative profits. Suppose $\varphi(a_1,a_2)$ denotes the "efficiency" distribution, or the number of potential production units times the density of production capabilities (a_1,a_2). Total production and total input usage is determined as

$$Q = \int_{\mathcal{A}(w/p,r/p)} \varphi(a_1,a_2)\, da_1\, da_2$$

$$\text{(4.44)} \qquad L = \int_{\mathcal{A}(w/p,r/p)} a_1\varphi(a_1,a_2)\, da_1\, da_2$$

$$K = \int_{\mathcal{A}(w/p,r/p)} a_2\varphi(a_1,a_2)\, da_1\, da_2$$

The primitive feature of this model is the efficiency distribution, and an induced aggregate production relation $Q = S(L,K)$ can result from solving out the above system, eliminating w/p and r/p. Houthakker originally noted that if φ is a Pareto distribution, or $\varphi(a_1,a_2) = Aa_1^{\alpha_1-1}a_2^{\alpha_2-1}$, then the induced aggregate production relation is in Cobb Douglas form $Q = C\, L^{\beta_1}K^{\beta_2}$, where $\beta_1 = \alpha_1/(\alpha_1+\alpha_2+1)$ and $\beta_2 = \alpha_2/(\alpha_1+\alpha_2+1)$.

This kind of model has been developed by numerous authors, most extensively by Kazuo Sato (1975) and Leif Johansen (1972). In terms of recent empirical implementations, Hildenbrand (1981) employs a model of this kind for Norwegian tanker production, where he characterizes the efficiency distribution directly from individual firm data. Finally, Heckman and V.K. Chetty (1986) extend the basic model to include an adjustment equation for capital over time, and apply it to the analysis of U.S. manufacturing. While these models are interesting alternatives to standard continuous production models, their applicability hinges on the strong assumption of limited input substitutability at the level of individual production units, as well any assumed shape and evolution of the efficiency distribution over time.

Discreteness of individual reactions also plays a central role in some recent models of macroeconomic adjustment. A primary example is the (s,S) model of aggregate inventory dynamics developed by Caplin and Daniel Spulber (1987) and Ricardo Caballero and Eduardo Engel (1991,1992). Here, discreteness arises because individual firms adjust inventories according to threshold criterion - firm i waits until its inventory reaches level s_i, at which point the inventory is increased to S_i. Aggregate adjustment occurs sluggishly as different firms react to shocks at different times. The distribution of reactions provide the central structure of these models.

Sluggishness in aggregate investment responses also arise from irreversibilities of investment decisions by individual firms. Guiseppe Bertola and Caballero (1990) give a detailed analysis of the dynamic aggregate behavior of an economy populated by agents behaving according to (s,S) rules.

The empirical implementation of these adjustment models is in an early stage of development. In particular, the initial efforts have been focused solely on broad aggregate implications, and studied using aggregate data series alone. The potentially realistic features of the adjustment processes in these models need to be verified using data from individual firms, and methods developed for tracking the sectoral composition of aggregate inventory or investment statistics.[31]

4.6 Recent Work on Microsimulation

A primary reason for studying aggregation over individual agents is for simplification in economic modeling. A single, properly specified equation for an economic aggregate is useful in a larger model of the economy as a method of summarizing the behavior of a large group of agents, be they producers or consumers. We now turn to a brief discussion of work that models heterogeneous agents explicitly, without concern for whether a parsimonious aggregate model can be formulated.

One emerging trend in macroeconomic research is the study of model economies with two or three different (kinds of) consumers or other agents. The purpose of this work is to look in detail at heterogeneity in the context of markets for risk sharing, where such markets are either efficient or in some way incomplete. Recent work of this kind includes the two-agent models of Bernard Dumas (1989) and of John Heaton and Deborah Lucas (1992), who also discuss references to this recent literature. It is clear that with two or three agents, this kind work is unlikely to give a realistic depiction of heterogeneity as it exists in a real world economy, and therefore has limited applicability to practical questions. However, the superficial treatment of heterogeneity facilitates another purpose, which is to address difficult questions on the workings of markets for risk sharing. Consequently, this work may yield valuable insights on the interplay between market interactions and differences between agents.

More germane to our discussion are full scale microsimulation models. As discussed in the opening remarks, it is difficult to argue against the microsimulation approach for modeling aggregates on logical grounds. We have stressed how it is essential to model individual behavior, and it is a natural next step to conclude that individual modeling should be carried out without any additional considerations, such as whether the purposes of aggregate prediction are served. After all, one can add up across models of individual behavior, giving aggregate responses that are behaviorally based.

Logical correctness, however, does not translate to practical tractability. Even with a small number of variables representing individual heterogeneity, an extensive setup is required for a full implementation of microsimulation: a complete model of individual behavior linked to a model of the evolution of heterogeneous individual attributes. For instance, a general model of household spending behavior must be linked to a model of the evolution of the demographic structure of the population, let alone a model of wage and income determination. As demonstrated by Cowing and McFadden (1984), the complexities inherent in this process preclude validation of aggregate results from such a model relative to more parsimonious modeling of aggregate data patterns. Our discussion of models that account for aggregation has focused on how the required inputs for applications can be summarized in modeling aggregate data patterns.[32]

As developments in computational power progress unabated, it is natural to expect that methods of implementing and validating microsimulation models will be developed in the future. As part of our survey, we discuss two recent types of work that overcome the existing shortcomings of microsimulation in different ways. The first is the Joint Committee on Taxation's (1992) model of forecasting the impacts of tax policy changes. This model follows in the tradition of tax policy models developed by the NBER (see Daniel Feenberg and Harvey Rosen (1983), among others). Impacts of changes in tax policy are simulated at the individual level by recomputing individual tax forms, combined with some assumptions on reporting differences and other behavioral changes induced by the policy changes. This model is likely the most important microsimulation model in use today, as it is the primary source of estimates for tax policy changes for the U.S. Congress, and requests for its results have grown dramatically in recent years (from 348 in 1985 to 1,460 in 1991, for instance).

The simplification employed by the Joint Committee on Taxation's model is aptly described by a section heading of their 1992 report, "Holding Fixed the Level of Macroeconomic Aggregates." In particular, the model holds constant the effects of economic growth, monetary policy and other changes in fiscal policy, and focuses solely on the distributional impacts of tax policy changes. By removing the effects of interest rates and price level (inflation) changes, the projection of tax impacts for individuals is greatly simplified. However, this feature places a large proviso on forecasts from the model. Comparisons between the results of this model and results from less detailed macroeconomic models (that study tax effects together with the effects of macroeconomic aggregates) are likewise somewhat problematic.

The second kind of microsimulation model is described in Heckman and Walker (1989,1990), who give the results of a full scale "horse race" between a fully nonlinear microsimulation model and simple aggregate forecasting equations. The object of this study is the forecasting of fertility rates, and the comparison is between simple time series models of aggregate fertility rates and the results of simulating a dynamic individual model of durations between births. The net result is that the microsimulation model out performs the simple forecasting equations along several criteria. While this model is too complicated to discuss in any detail here, these results raise hopes that microsimulation methods may be profitably applied to forecasting aggregate data. An interesting feature of the model is that the inherent dynamics serve to simplify the difficulties in creating inputs for the microsimulation. In particular, the dynamic features of the model (durations between births are determined by previous birth history) create the distributions required for predicting future fertility endogenously.

5. Some Conclusions

One of the most difficult problems faced by beginning readers of the literature on aggregation over individuals is its own heterogeneity; it consists of a wide range of seemingly unrelated problems and approaches. As such, we have only given brief glimpses at pieces of a rapidly growing area. Our approach to surveying recent developments was to spell out conceptual issues for interpreting equations estimated with aggregate data, and then discuss specific approaches with the interpretation issues in mind. This posture was chosen as a way of focusing attention on the properties valuable for empirical applications, which is the most natural future avenue for progress.

In many ways the most important development of the work of the last decade is the demonstration of how individual heterogeneity can actually be incorporated in the modeling of aggregate data. While the models we have discussed are often simple, and many unsolved questions remain for accommodating more complicated models (including market interactions), the "aggregation problem" is no longer a mysterious proviso of macroeconomic data analysis, to be given lip service and then ignored. The issues we have discussed concerning the relative importance of individual heterogeneity and aggregate dynamics certainly suggest that the most valuable applied work in this area is yet to come.

A note on the historical setting of this work is useful to place it in context. In particular, the work we have surveyed can be regarded as attempts to merge two separate trends in research. The first is empirical macroeconomics, which has evolved through the development of exceedingly sophisticated behavioral models, and applied either formally or informally through the guise of a representative agent. The value of this work lies in how it has permitted empirical measurement to be focused on specific, easily understood issues. Representative agent models were first used for interpretable measurement of substitution patterns in consumption and production, and has proceeded through the demonstration of how primitive structure (technology and preferences) relates to observed choices by individuals under uncertainty.

The second trend involves theoretical work devoted to the implications of heterogeneity over individuals. This work created an increasingly dismal view of representative agent modeling, by showing that heterogeneity could be neglected only in very restricted, unrealistic settings. The strongest form of criticism of empirical aggregate data modeling came in the work of Gerard Debreu (1974) and Hugo Sonnenschein (1972), as surveyed by Wayne Schafer and Sonnenschein (1982), that stated that no restrictions on aggregate excess demands could be adduced from economic logic alone, aside from Walras Law and lack of money illusion. In particular, they demonstrated that one can begin with any formula with those properties, and construct an economy with that formula as the aggregate excess demand function. The Debreu-Sonnenschein work was interpreted by most as stating that, because no specific restrictions on aggregate relationships were guaranteed, there was no rationale for structuring models to be consistent with a representative agent. Representative agent models can never have a firm foundation from economic theory alone.

While true, this interpretation is purely negative, and does not suggest productive directions for empirical work. A more constructive interpretation of the Debreu-Sonnenschein work is it points out the need to add more structure to justify aggregate data models. In particular, to study aggregate data from the U.S. economy, what is relevant are characteristics of the U.S. economy. Relative to empirical economics, who cares if an artificial general equilibrium model could be constructed for any aggregate data pattern? What needs to be studied are actual observed aggregate data patterns as they are related to the actual characteristics of the U.S. economy. To the extent that economic theory is valuable for giving structure to individual behavior, it should be applied at the individual level, and there is nothing wrong with tracing the aggregate implications of such behavior. The Debreu-Sonnenschein results are more appropriately applied to methods that are not based on observed data series, because the criticism states that one can make up a model to generate any arbitrary excess demand structure, and therefore generate any answers one wants.

The work we have surveyed can be seen as the initial attempts to build empirical models that are applicable to the applied questions of aggregate data, but retain the feature of modeling behavior at the individual level. Because of the bridge between micro and macro levels in these models, structure from individual level decision is brought to bear on aggregate data patterns in a consistent way. This linkage permits behavioral responses to be studied with aggregate data, and future aggregate data patterns to be simulated in a fashion consistent with the heterogeneous composition of the population.

There is a broad set of prescriptions for empirical modeling available from the work we have discussed. First, in constructing models that measure aspects of behavior, one must begin "from the ground up," or always begin with a model of behavior at the individual level. There is no sufficiently broad or realistic scenario in which one can begin with a representative agent's equations without explicitly considering the impact of heterogeneity. Whether a representative agent model fits the data or not, there is no realistic paradigm where the parameters of such a model reflect only behavioral effects, uncontaminated by compositional considerations. The application of restrictions appropriate for individual behavior directly to aggregate data is a practice without any foundation, and leads to biases that are impossible to trace or measure with aggregate data alone. The only way individual level restrictions can be consistently applied to aggregate data is through the linkage provided by an assumed aggregation structure.

To implement a consistently constructed model of individual behavior and aggregate data, it is important to stress that all relevant data should be employed. In this context, this means that an aggregate level model is applied to aggregate level data, the individual level model is applied to individual level data, and consistently derived equations are applied to partially disaggregated data, such as those on coarse groupings of the population. All types of data are relevant to a single model, or measurement of a single set of behavioral parameters.

We have glossed over the potential data problems of comparing individual and aggregate level data, because of the overall importance of modeling at all levels simultaneously. In particular, the potential for measurement problems in individual level data does not give proper excuse for ignoring the necessary connections between individual behavior and aggregate data. When one suspects problems of conceptual incompatibility, a more informative approach is to check for the implications of such problems within the context of a fully consistent individual-aggregate level model. For instance, one way of judging a "cure" for a measurement problem in individual level data is to see if the resulting parameter estimates are comparable with those obtained from aggregate data.

Hand-in-hand with the necessity of using all relevant data is the necessity of checking or testing all relevant assumptions underlying a model.

Aside from a platitude of good empirical work, it is important to stress the testing aspect here because altogether too little attention has been paid to checking or testing assumptions required for aggregation, relative to assumptions on the form of individual behavioral models. For instance, exact aggregation models rely on intrinsic linearity in the form of individual level equations, and testing such restrictions should take on a high priority, and be implemented with individual level data. For aggregation over intrinsically nonlinear models, specific distributional assumptions are required, and likewise become testable implications of the model. In essence, one should avoid the temptation to regard the aggregation structure as secondary to the individual model of economic behavior, focusing on one and ignoring the other, as both types of structure have equal bearing on the subsequent model of aggregate data. A fact of life is that only the crudest implications of heterogeneity can be studied with aggregate data alone - while distributional data should be included in any study of aggregate data to check the specification of estimated macroeconomic equations, the most informative assessments of aggregation structure will come from studying individual level cross section or panel data.

In line with this is a cautionary remark about the natural temptation to just create a "story" to "justify" common reduced form or representative agent models. The problem is that for any equation connecting aggregates, there are a plethora of behaviorally different "stories" that could generate the equation, which are observationally equivalent from the vantage point of aggregate data alone. If one invents a paradigm that is not consistent with individual data, or based on fictitious coordination between agents, then the results of estimating an aggregate equation based on that paradigm are not well founded, and are not to be taken seriously. For an arbitrary example, suppose that one applies a representative agent model of commodity demands, asserting the existence of community indifference curves through optimal redistribution of income in each time period, following Samuelson (1956). The results of such estimation have credibility only if one can find convincing evidence that such redistribution is actually occurring, and occurring to the extent required for Samuelson's result (or maintaining constant marginal social welfare for each income level). Checking a "story" that motivates an aggregate data model *always* requires looking beyond aggregate data to the underlying process.

With these prescriptions, one should be quite optimistic about the overall prospects for dealing with the problem of aggregation over individuals, or understanding the implications of individual heterogeneity in macroeconomic analysis. Approaches that neglect individual heterogeneity, such as pure representative agent modeling, should be abandoned. However, there is no reason why the wide array of individual behavioral models developed under the representative agent paradigm cannot be applied at the individual level, and used as a consistent foundation for studying macroeconomic data.

Notes

[1] It is useful to stress that our basic concerns with "individual heterogeneity" refer to differences between groups that are observable in the context of an econometric analysis of individual level data, and not arbitrarily fine distinctions that could characterize each individual differently. For illustration, suppose first that you work for a large corporation, and your job is to assess the authenticity of claims for disability payments. Sam Jones has filed a claim, because of back pain that he attributes to a fall he took while at work. In this instance, your job is to decide on a fine distinction of individual heterogeneity; namely, whether Mr. Jones is actually unable to work. Alternatively, suppose that you are the executive in charge of forecasting the costs of future disability claims to the company. In this case, you do not particularly care about whether Mr. Jones is disabled, but you do care about what percentage of workers typically become disabled, relative to age, skill level and type of job. It is this latter notion of individual heterogeneity that is germane to our discussion of modeling economy-wide aggregates. This distinction also seems to underly Kirman's (1992) puzzling remarks associating representative agent models and exact aggregation models.

[2] Further, we do not address the issues raised by endogeneity of individual differences; for instance, the potential impact of endogeneity of family size in studies of demand. Some topics we cover (such as econometric estimation) can accommodate endogeneity with few modifications, however others (such as aggregation with nonlinear individual models) involve many more complications.

[3]For instance, Section 4.4 includes coverage of Euler equation estimation in intertemporal consumption models. There y_{it} can denote current consumption, M_{it} lagged consumption, p_t interest rates and common information sets, and A_{it} demographic differences, innovations, etc.

[4]While the features of this theory are discussed by Jacob Marshak (1939) and Pieter DeWolff (1941), the main contributions date from Gorman (1953) and Henri Theil (1954), through John Muellbauer (1975,1976), Lawrence Lau (1977, 1982), Dale Jorgenson, Lau and Stoker (1982) and John Heineke and Herschel Sheffrin (1990), among many others.

[5] This is not a loaded phrase indicating many detailed steps. In particular, our assumption that M is lognormal states that $(\ln M - \mu_t)/\Sigma_t$ is standard normal (with mean 0 and variance 1). Therefore, we divide both sides of (3.18) by β_2, subtract ln M, add μ_t, and finally divide by Σ_t. This gives (3.19), which is in a convenient form for solving for the aggregate buying percentage $E_t(y)$.

[6]Related issues are addressed by Harry Kelejian (1980).

[7]Equation (3.25) suggests we have returned to the case where y_{it} is continuous, but all conclusions are valid if (3.25) holds with y_{it}, x_{1it}, x_{2it} taking on discrete or otherwise limited values.

[8] A numerical example of the trending phenomena is given in Stoker (1984a).

[9] All remarks would apply if the distribution of ϵ varied over time in a known way.

[10] For example, suppose that the discrete choice model (3.16) included a normal disturbance: $y_{it} = 1$ if $1 + \beta_1 \ln p_t + \beta_2 \ln M_{it} + \epsilon \geq 0$, $y_{it} = 0$ otherwise, where ϵ is normal with mean 0 and variance σ_0^2. Our notation here has p and ϵ as above, but x = M. The conditional expectation $E(y|p,M)$ is the percentage of people at income level M who buy the product when price is p; here $E(y|p,M) = \Phi[\sigma_0^{-1}(1 + \beta_1 \ln p + \beta_2 \ln M)]$. With $\rho(M|\mu_t,\Sigma_t)$ the lognormal distribution, the aggregate model is given by (3.20) where $\beta_2\Sigma_t$, the denominator inside the brackets, is replaced by $(\beta_2^2\Sigma_t^2 + \sigma_0^2)^{1/2}$, which permits recoverability of β_1, β_2 and σ_0 from aggregate data.

[11] Models (3.38) and (3.39) can easily be formulated via orthogonality conditions, with estimation carried out using instrumental variables or another generalized method of moments method. This would accommodate various kinds of endogenous predictor variables, as well as the standard setup for models of behavior under uncertainty (Section 4.4). This proviso applies throughout our discussion below, where we discuss least squares estimation for simplicity.

[12] These issues arise in cohort analysis, and the use of repeated cross sections for estimation of dynamic models; see Angus Deaton (1985), among others. While we do not delve into the differences between using panel data and repeated cross sections here, it is likely that Robert Moffitt's (1991) arguments in favor of using repeated cross sections (avoiding the attrition that plagues long panels) have some validity here.

[13] To the extent that the disturbance in the aggregate equation is the average of the disturbance in the individual equations, then these variances would reflect grouping or size heteroskedasticity as well. In this regard, one might ask whether more efficient estimates are available by taking into account the correlation between the individual and aggregate data; while the answer is yes, when the cross section is small relative to the population, the adjustments required (with random errors) are negligible (c.f. Stoker (1977)). For instance, for family expenditure data in the United States, one might observe 10,000 households in a cross section, with a population size of N_t = 90 million.

[14] These connections are developed in Stoker (1982, 1986a, 1986d). Stoker (1985) discusses similar connections for measuring the sensitivity of aggregate forecasts to individual parameter estimates.

[15] This connection follows from integration-by-parts, or the "generalized information matrix equality" familiar to readers of econometrics. In particular, $\partial E[g(x)]/\delta\mu = \mathrm{Cov}[\ell(x),g(x)]$ where $\ell(x) = \partial \ln \rho(x|\mu)/\partial\mu$ (provided boundary terms vanish). $\hat{d}$ estimates $[\mathrm{Cov}(\ell,x)]^{-1}[\mathrm{Cov}(\ell,y)] = [\partial E(x)/\partial\mu]^{-1}\partial E(y)/\partial\mu = \partial E(y)/\partial E(x)$ at time t.

[16] Stoker (1986c) develops this structure somewhat differently, based on stable demand behavior within each range of the income distribution. This posture facilitates measures of the extent to which individual nonlinearity "averages away" in aggregate data.

[17] A further study of note is the use of the age distribution to explain housing prices by Gregory Mankiw and David Weil (1989), although these authors curiously omit income and other economic variables that would seem appropriate for housing demand.

[18] Theil (1975) lays out the original foundation for per-capita application of the Rotterdam demand model, with more general formulations given in William Barnett (1979) among others.

[19] Demand analysis provides perhaps the only area where the theoretical implications on income structure of exact aggregation models are well understood. In particular, Gorman (1981) studies the demand $q(p,M) = \sum b_j(p)\ \psi_j(M)$, where the $\psi_j(M)$ terms could be of any form. In a remarkable analysis, he shows that there can be at most three linearly independent $\psi_j(M)$ terms, which is referred to as the "Gorman Rank 3" result. He further characterizes the admissible $\psi_j(M)$ terms; including the power and log terms used in the models we discuss later, as well as trigonometric terms. Lewbel (1991) discusses his extensions of these results in the context of demand rank, which unifies this work with exact aggregation, and work on generalized Slutsky conditions of Erwin Diewert (1977) and Stoker (1984b).

[20]These models arise from the path breaking work of John Muellbauer (1975) on when aggregate demand depends on "representative income," and adopt Engel curve formulations proposed earlier by Harold Working (1943) and Conrad Leser (1960).

[21] An important example that we have not covered is the Quadratic Expenditure System of Pollak and Wales (1979). This model is quadratic in total expenditure, which results in aggregate demand depending on $\bar{M}_t$, $\bar{M}_t^2$ and the variance $V_t = N_t^{-1} \sum (M_{it} - \bar{M}_t)^2$. All the same remarks apply here - for instance, to implement a model based on this system with aggregate data requires either observing V_t or adopting a restriction relating V_t to $\bar{M}_t$. From the development in Pollak and Wales (1992), it is clear that the QES can provide the basis for an aggregate model that allows recoverability, including accounting for demographic differences across families.

[22]It is interesting to note how a parallel development is underway in general equilibrium theory, in part due to the increasing recognition that the famed Arrow-Debreu model is vacuous for purposes of empirical work. For instance, Xavier Freixas and Andreu Mas-Collel (1987) derive very similar forms to those of Muellbauer (1975) by studying aggregate revealed preference properties. In line with the models we have just discussed, Werner Hildenbrand (1983, 1992) proposes using nonparametric methods for introducing income distribution information in studying aggregate income effects, with estimates given by Wolfgang Härdle, Hildenbrand and Michael Jerison (1991). Jean-Michel Grandmont (1992) studies the introduction of demographic differences through family equivalent scales, obtaining results with some similarity to those of Gary Becker's (1962) random demand model.

[23] Jorgenson, Daniel Slesnick and Stoker (1988) give another example of an estimated exact aggregation model.

[24] Related applications of this kind can be found in Terence Barker and Pesaran (1990), among others. The discussion of random coefficients likewise applies to the traditional justification for the Rotterdam model of demand, c.f. Theil (1975) and Barnett (1979), among many others.

[25] The main situation where a primary focus on aggregate data can be justified is with a linear model where the individual predictor variables are observed up to (tightly structured) errors; the measurement errors in the individual data have to have mean 0 across individuals, and be uncorrelated with individual marginal effects. In this case, regression with individual data involves the standard bias, but the average of the observed predictors closely matches the average of the true predictors (the predictors with error have the same "common factor" as the true predictors). See Dennis Aigner and Stephen Goldfeld (1974), among others.

[26] In the notation of Section 3.1, we have $y_{it} = C_{it}$ and $x_{it} = (A_{it}C_{it-1}, (1-A_{it})I_{it})$.

[27] The notation of Section 3 coincides here as $y_{it} = C_{it}$, $M_{it} = C_{it-1}$ and A_{it} contains v_{it}. Common information would coincide with p_t, and household specific information would enter through A_{it}.

[28] For instance, a consumption model of the kind in (4.39, 4.40) could be used as part of a real business cycle simulation model, such as those discussed by Edward Prescott (1986). If such equations are estimated using individual data, their use represents a more scientific method of calibrating such a model to micro data than is often practiced, such as setting approximate parameter values for a representative agent from older studies of micro data, and matching factor shares and other aggregates.

[29] Our discussion has not focused on situations where prices vary across individuals. In such cases, the varying prices are treated like other varying attributes, and restrictions to accommodate such variation are combined with restrictions on price effects from the basic choice model. For example, Muellbauer (1981) uses an exact aggregation model to study labor supply with varying wages across individuals.

[30] Aggregate simulations of discrete choice models are given by Colin Cameron (1990), as well as many references to the transportation economics literature. The recent literature on monopolistic competition contains theoretical analysis of aggregation and discrete response models, such as Egbert Dierker (1989) and Andrew Caplin and Barry Nalebuff (1991), although these ideas have not been developed for empirical study. Other work of note concerns the aggregation structure of discrete response models germane to marketing; see Simon Anderson, André de Palma and Jacques-Francois Thisse (1989), for instance, as well as Greg Allenby and Peter Rossi's (1991) study of why macro "logit" models demonstrate good fitting properties to aggregate market shares.

[31] Pok-Sang Lam (1991) reports on the results of applying an (s,S) model to automobile data.

[32] Inputs that are endogenous generally make prediction more difficult, but are especially onerous in microsimulation models, because the entire distribution of endogenous inputs must be simulated. For instance, suppose that one was interested in forecasting welfare payments, and at issue was whether the welfare system actually induced families on welfare to have more children. In this case, family size could not be treated as exogenous. To simulate the effect of a policy change in welfare payments, induced changes in the distribution of family size would have to be simulated, which is much more complicated than projecting exogenous changes in family size distribution.

References

Aigner, Dennis J. and Goldfeld, Stephen M. (1974), "Estimation and Prediction from Aggregate Data When Aggregates are Measured More Accurately that Their Components," *Econometrica*, 42, 113-34.

Allenby, Greg M. and Rossi, Peter E. (1991), "There is No Aggregation Bias: Why Macro Logit Models Work," *Journal of Economic and Business Statistics*, 9, 1-14.

Anderson, Simon P.; de Palma, André and Thisse, Jacques-Francios (1989), "Demand for Differentiated Products, Discrete Choice Models and the Characteristics Approach," *Review of Economic Studies*, 56, 21-36.

Barker, Terence S. and Pesaran, M. Hashem (1990), *Disaggregation in Econometric Modeling*, London, Routledge.

Barnett, William A. (1979), "Theoretical Foundations of the Rotterdam Model," *Review of Economic Studies*, 46, 109-130.

Berndt, Ernst R.; Darrough, Masako N. and Diewert W. Erwin (1977), "Flexible Functional Forms and Expenditure Distributions: An Application to Canadian Consumer Demand Functions," *International Economic Review*, 18, 651-675.

Becker, Gary S. (1962), "Irrational Behavior and Economic Theory," *Journal of Political Economy*, 70, 1-13.

Bertola, Guiseppe and Caballero, Ricardo J. (1990), "Irreversibility and Aggregate Investment," Columbia Working Paper, #488.

Blackorby, Charles; Boyce, Richard and Russell, Robert R. (1978), "Estimation of Demand Systems Generated by the Gorman Polar Form: A Generalization of the S-Branch Utility Tree," *Econometrica*, 46, 345-364.

Blackorby, Charles; Primont, Daniel and Russell, Robert R. (1978), *Duality, Separability and Functional Structure: Theory and Economic Applications*, Amsterdam, North Holland.

Blinder, Alan S. (1975), "Distribution Effects and the Aggregate Consumption Function," *Journal of Political Economy*, 83, 447-75.

Blundell, Richard (1988), "Consumer Behavior: Theory and Empirical Evidence - A Survey," *Economic Journal*, 98, 16-64.

Blundell, Richard; Pashardes, Panos and Weber, Guglielmo (1992), "What Do We Learn About Consumer Demand Patterns from Micro Data," draft, University College London, forthcoming *American Economic Review*.

Browning, Martin (1992), "Children and Household Economic Behavior," *Journal of Economic Literature*, 30, 1434-1475.

Buse, Adolph (1992). "Aggregation, Distribution and Dynamics in the Linear and Quadratic Expenditure Systems," *Review of Economics and Statistics*, 74, 45-53.

Caballero, Ricardo J. and Engel, Eduardo M.R.A. (1991), "Dynamic (S,s) Economies," *Econometrica*, 59, 1659-1686.

Caballero, Ricardo J. and Engel, Eduardo M.R.A. (1992), "Microeconomic Adjustment Hazards and Aggregate Demands," draft, MIT Department of Economics.

Cameron, A. Colin (1990), "Aggregation in Discrete Choice Models: An Illustration of Nonlinear Aggregation," in T.S. Barker and M.H. Pesaran, eds., *Disaggregation in Econometric Modeling*, London, Routledge.

Caplin, Andrew S. and Nalebuff, Barry (1991), "Aggregation and Imperfect Competition: On the Existence of Equilibrium," *Econometrica*, 59, 1-24.

Caplin, Andrew S. and Spulber, Daniel (1987), "Menu Costs and the Neutrality of Money," *Quarterly Journal of Economics*, 102, 703-726.

Chiappori, Pierre-André (1988), "Rational Household Labor Supply," *Econometrica*, 56, 63-90.

Chetty, V.K. and Heckman, James J. (1986), "A Dynamic Model of Aggregate Output Supply, Factor Demand and Entry and Exit for a Competitive Industry with Heterogeneous Plants," *Journal of Econometrics*, 33, 237-262.

Cowing, Thomas G. and McFadden, Daniel L. (1984), *Microeconomic Modeling and Policy Analysis*, New York, Academic Press.

Deaton, Angus S. (1985), "Panel Data from Time Series of Cross Sections," *Journal of Econometrics*, 30. 109-126.

Deaton, Angus S. and Muellbauer, John (1980a), "An Almost Ideal Demand System," *American Economic Review*, 70, 312-326.

Deaton, Angus S. and Muellbauer, John (1980b), *Economics and Consumer Behavior*, Cambridge, Cambridge University Press.

Debreu, Gerard (1974), "Excess Demand Functions," *Journal of Mathematical Economics*, 1, 15-23.

Dierker, Egbert, "Competition for Customers," Universität Bonn, Discussion Paper A-244.

Diewert, W. Erwin (1977), "Generalized Slutsky Conditions for Aggregate Demand Functions, *Journal of Economic Theory*, 15, 353-362.

Dumas, Bernard (1989), "Two-Person Dynamic Equilibrium in the Capital Market," *The Review of Financial Studies*, 2, 159-188.

Eichenbaum, Martin S.; Hansen, Lars Peter and Richard, Scott F. (1987), "Aggregate Durable Goods and Nonseparable Preferences in an Equilibrium Asset Pricing Model," NORC Discussion Paper 87-9.

Fair, Ray C. and Dominguez, Katherine M. (1991), "Effects of the Changing U.S. Age Distribution on Macroeconomic Equations," *American Economic Review*, 81, 1276-1294.

Feenberg, Daniel R. and Rosen, Harvey S. (1983), "Alternative Tax Treatment of the Family: Simulation Methodology and Results," Chapter 1 of M.S. Feldstein, ed., *Behavioral Simulation Methods in Tax Policy Analysis*, Chicago, University of Chicago Press.

Freixas, Xavier and Mas-Colell, Andreu (1987), "Engel Curves Leading to the Weak Axiom in the Aggregate," *Econometrica*, 55, 515-532.

Grandmont, Jean-Michel (1992), "Transformations of the Commodity Space: Behavioral Heterogeneity and the Aggregation Problem," *Journal of Economic Theory*, 57, 1-35.

Granger, Clive W.F. (1980), "Long-Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics*, 14, 227-38.

Granger, Clive W.F. (1987), "Implications of Aggregation with Common Factors," *Econometric Theory*, 3, 208-222.

Granger, Clive W.F.(1990), "Aggregation of Time Series Variables: A Survey," in Barker, Terence S. and Pesaran, M. Hashem, eds., *Disaggregation in Econometric Modeling*, London, Routledge.

Gorman, William M. (Terence) (1953), "Community Preference Fields," *Econometrica*, 21, 63-80.

Gorman, William M. (Terence) (1981), "Some Engel Curves," in Deaton, Angus S., ed. *Essays in the Theory and Measurement of Consumer Behavior*, Cambridge, Cambridge University Press.

Grunfeld, Yehuda and Griliches, Zvi (1960), "Is Aggregation Necessarily Bad?" *Review of Economics and Statistics*, 42, 1-13.

Hall, Robert E. (1978), "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 86, 971-987.

Hansen, Gary D. (1985), "Indivisible Labor and the Business Cycle," *Journal of Monetary Economics*, 16, 309-27.

Hansen, Lars Peter and Singleton, Kenneth J. (1983) "Stochastic Consumption, Risk Aversion and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 91, 249-65.

Härdle, Wolfgang; Hildenbrand, Werner and Jerison, Michael (1991), "Empirical Evidence for the Law of Demand," *Econometrica*, 59, 1525-1550.

Hausman, Jerry A.; Newey, Whitney K. and Powell, James L. (1992), "Errors-in-Variables Estimation of Some Engel Curves," draft, forthcoming *Journal of Econometrics*.

Heaton, John and Lucas, Deborah (1992), "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing," Sloan School of Management Working Paper.

Heckman, James, J. and Chetty, V.K. (1986), "A Dynamic Model of Aggregate Output, Factor Demand and Entry and Exit for a Competitive Industry with Heterogeneous Plants," *Journal of Econometrics*, 33, 237-262.

Heckman, James J. and Sedlacek, Guilherme (1985), "Heterogeneity, Aggregation and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," *Journal of Political Economy*, 93, 1077-1125.

Heckman, James J. and Walker, James R. (1989), "Forecasting Aggregate Period-Specific Birth Rates: The Time Series Properties of a Microdynamic Neoclassical Model of Fertility," *Journal of the American Statistical Association*, 84, 958-965.

Heckman, James J. and Walker, James R. (1990), "The Relationship Between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data," *Econometrica*, 58, 1411-1442.

Heineke, John M. and Shefrin, Herschel M. (1990), "Aggregation and Identification in Consumer Demand Systems," *Journal of Econometrics*, 44, 377-390.

Hildenbrand, Werner (1981), "Short Run Production Functions Based on Microdata," *Econometrica*, 49, 1095-1126.

Hildenbrand, Werner (1983), "On the Law of Demand," *Econometrica*, 51, 997-1019.

Hildenbrand, Werner (1992), "Market Demand, Theory and Empirical Evidence," Discussion Paper A-359, Universität Bonn.

Houthakker, Hendrick S. (1955), "The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis," 23, 27-31.

Johansen, Leif (1972), *Production Functions*, Amsterdam, North Holland.

Joint Committee on Taxation (1992), *Discussion of Revenue Estimation and Process*, JCS 14-92, August 13, 1992, U.S. Government Printing Office, Washington, D.C.

Jorgenson, Dale W.; Lau, Lawrence J. and Stoker, Thomas M. (1980), "Welfare Comparisons and Exact Aggregation," *American Economic Review*, 70, 268-272.

Jorgenson, Dale W.; Lau, Lawrence J. and Stoker, Thomas M. (1982), "The Transcendental Logarithmic Model of Aggregate Consumer Behavior," in Basmann, Robert L. and Rhodes, George, eds., *Advances in Econometrics, Vol. 1*, Greenwich, JAI Press, 97-238.

Jorgenson, Dale W. and Slesnick, Daniel T. (1984), "Aggregate Consumer Behavior and the Measurement of Inequality," *Review of Economic Studies*, 51, 369-392.

Jorgenson, Dale W.; Slesnick, Daniel T. and Stoker, Thomas M. (1988), "Two Stage Budgeting and Exact Aggregation," *Journal of Business and Economic Statistics*, 6, 313-325.

Jorgenson, Dale W. and Stoker, Thomas M. (1986), "Nonlinear Three Stage Least Squares Pooling of Cross Section and Average Time Series Data," in Moroney, John R., ed., *Advances in Statistical Analysis and Statistical Computing, Vol. 1*, Greenwich, JAI Press, 87-116.

Kelejian, Harry J. (1980), "Aggregation and Disaggregation of Nonlinear Equations," in Jan Kmenta and James B. Ramsey, eds., *Evaluation of Econometric Models*, New York, Academic Press.

Kirman, Alan P. (1992), "Whom or What Does the Representative Individual Represent," *Journal of Economic Perspectives*, 6, 117-136.

Lam, Pok-Sang (1992), "Permanent Income, Liquidity, and Adjustments of Automobile Stocks: Evidence from Panel Data," *Quarterly Journal of Economics*, 106, 203-230.

Lau, Lawrence J. (1977), "Existence Conditions for Aggregate Demand Functions," Institute for Mathematical Studies in the Social Sciences, Stanford University, Technical Report No. 248.

Lau, Lawrence J. (1982), "A Note on the Fundamental Theorem of Exact Aggregation," *Economic Letters*, 9. 119-126.

Lee, K.; Pesaran, M. Hashem and Pierse, Richard G. (1990), "Testing for Aggregation Bias in Linear Models," *Economic Journal*, 100, 137-150

Leser, Conrad E.V. (1963), "Forms of Engel Functions," *Econometrica*, 31, 694-703.

Lewbel, Arthur (1989a), "A Demand System Rank Theorem," *Econometrica*, 57, 701-706.

Lewbel, Arthur (1989b), "Exact Aggregation and a Representative Consumer," *Quarterly Journal of Economics*, 104, 622-33.

Lewbel, Arthur (1991), "The Rank of Demand Systems: Theory and Nonparametric Estimation," *Econometrica*, 59, 711-730.

Lewbel, Arthur (1992), "Aggregation and Simple Dynamics," draft, Brandeis University.

Lippi, Marco (1988), "On the Dynamic Shape of Aggregated Error Correction Models," *Journal of Economic Dynamics and Control*, 12, 561-585.

MaCurdy, Thomas E. (1987), "A Framework for Relating Microeconomic and Macroeconomic Evidence on Intertemporal Substitution," in Bewley, Truman F., ed., *Advances in Econometrics, Fifth World Congress, Vol. II*, 149-176.

Mankiw, N. Gregory and David N. Weil (1989), "The Baby Boom, the Baby Bust, and the Housing Market," *Regional Science and Urban Economics*, 19, 235-258.

Marshak, Jacob (1939), "Personal and Collective Budget Functions," *Review of Economic Statistics*, 21, 161-170.

Modigliani, Franco (1970), "The Life Cycle Hypothesis of Saving and Intercountry Differences in the Saving Ratio," in W. Eltis, M. Scott and I. Wolfe, eds. *Induction, Growth and Trade*, Oxford, Clarendon Press.

Moffitt, Robert (1991), "Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross Sections," Draft, Brown University.

Muellbauer, John (1975), "Aggregation, Income Distribution and Consumer Demand," *Review of Economic Studies*, 42, 525-543.

Muellbauer, John (1976), "Community Preferences and the Representative Consumer," *Econometrica*, 44, 979-1000.

Muellbauer, John (1981), "Linear Aggregation in Neoclassical Labor Supply," *Review of Economic Studies*, 48, 21-36.

Pesaran, M. Hashem; Pierce, Richard G. and Kumar, M.S. (1989), "Econometric Analysis of Aggregation in the Context of Linear Prediction Models," *Econometrica*, 57, 861-888.

Pollak, Robert A. (1985), "A Transactions Cost Approach to Families and Households," *Journal of Economic Literature*, 23, 581-608.

Powell, James L. and Stoker, Thomas M. (1985), "Estimation of Complete Aggregation Structures," *Journal of Econometrics*, 30, 317-344.

Prescott, Edward C. (1986), "Theory Ahead of Business Cycle Measurement," *Quarterly Review*, Federal Reserve Bank of Minneapolis, 10, 9-22.

Rubinstein, Mark, (1974), "An Aggregation Theorem for Securities Markets," *Journal of Financial Economics*, 1, 224-244.

Samuelson, Paul A. (1956), "Social Indifference Curves," *Quarterly Journal of Economics*, 73, 1-22.

Sato, Kazuo (1975), *Production Functions and Aggregation*, Amsterdam, North Holland.

Schafer, Wayne and Sonnenschein, Hugo (1982), "Market Demand and Excess Demand Functions," in Arrow, K.J. and M.D. Intriligator, eds, *Handbook of Mathematical Ecoonomics, Vol. II*, New York, North Holland.

Sonnenschein, Hugo (1972), "Market Excess Demand Functions," *Econometrica*, 40, 549-563.

Stoker, Thomas M. (1978), "The Pooling of Cross Section and Average Time Series Data," draft, Harvard University.

Stoker, Thomas M. (1982), "The Use of Cross Section Data to Characterize Macro Functions," *Journal of the American Statistical Association*, 77, 369-380.

Stoker, Thomas M. (1984a), "Completeness, Distribution Restrictions and the Form of Aggregate Functions," *Econometrica*, 52, 887-907.

Stoker, Thomas M. (1984b), "Exact Aggregation and Generalized Slutsky Conditions," *Journal of Economic Theory*, 33, 368-377.

Stoker, Thomas M. (1985), "Aggregation, Structural Change and Cross Section Estimation," *Journal of the American Statistical Association*, 80, 720-729.

Stoker, Thomas M. (1986a), "Aggregation, Efficiency and Cross Section Regression," *Econometrica*, 54, 177-188.

Stoker, Thomas M. (1986b), "The Distributional Welfare Impacts of Rising Prices in the United States," *American Economic Review*, 76, 335-349.

Stoker, Thomas M. (1986c), "Simple Tests of Distributional Effects on Macroeconomic Equations," *Journal of Political Economy*, 94, 763-795.

Stoker, Thomas M. (1986d), "Consistent Estimation of Scaled Coefficients," *Econometrica*, 54, 1461-1482.

Stoker, Thomas M. (1992), *Lectures on Semiparametric Econometrics*, CORE Lecture Series, Louvain-la-Neuve, CORE Foundation.

Theil, Henri (1954), *Linear Aggregation of Economic Relations*, Amsterdam, North Holland.

Theil, Henri (1975), *Theory and Measurement of Consumer Demand, Vol.1*, Amsterdam, North Holland.

Trivedi, Pravin K. (1985), "Distributed Lags, Aggregation and Compounding: Some Econometric Implications," *Review of Economic Studies*, 52, 19-35.

Van Daal, Jan and Merkies, Arnold H.Q.M. (1984), *Aggregation in Economic Research*, Dordrecht, D. Reidel.

de Wolff, Pieter (1941), "Income Elasticity of Demand, A Micro-Economic and a Macro-Economic Interpretation," *Economic Journal*, 51, 140-145.

Working, Harold (1943), "Statistical Laws of Family Expenditure," *Journal of the American Statistical Association*, 38, 43-56.

Zellner, Arnold (1969), "On the Aggregation Problem, a New Approach to a Troublesome Problem," in *Estimation and Risk Programming, Essays in Honor of Gerhard Tintner*, Berlin, Springer.