Chapter 68

MODELS OF AGGREGATE ECONOMIC RELATIONSHIPS THAT ACCOUNT FOR HETEROGENEITY*

RICHARD BLUNDELL
Institute for Fiscal Studies and Department of Economics, University College London, Gower Street, London, WC1E 6BT, UK

THOMAS M. STOKER
Sloan School of Management, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02138, USA

Contents

Abstract 4610
Keywords 4610
1. Introduction 4611
2. Consumer demand analysis 4615
   2.1. Aggregation of consumer demand relationships 4616
      2.1.1. Various approaches: Exact aggregation and distributional restrictions 4617
      2.1.2. Demand and budget share models 4619
      2.1.3. Aggregation in rank 2 and rank 3 models 4620
      2.1.4. Heterogeneous attributes 4622
   2.2. Empirical evidence and the specification of aggregate demand models 4623
      2.2.1. What do individual demands look like? 4623
      2.2.2. The implications for aggregate behavior 4625
   2.3. Aggregation of demand without individual structure 4628
3. Consumption and wealth 4629
   3.1. Consumption growth and income shocks 4630
      3.1.1. Idiosyncratic income variation and aggregate shocks 4631
      3.1.2. Income shocks and insurance 4632
      3.1.3. Incomplete information 4633
   3.2. Aggregate consumption growth with precautionary saving 4634

* A previous version of this chapter was published as “Heterogeneity and Aggregation” in the Journal of Economic Literature, vol. 43, 2, pp. 347–391. We are grateful for comments from numerous colleagues, especially Orazio Attanasio, Martin Browning, Jim Heckman, Arthur Lewbel and the reviewers. Thanks are also due to Zoë Oldfield and Howard Reed who helped organise the data for us. Financial support from the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at IFS is gratefully acknowledged. The usual disclaimer applies.

Handbook of Econometrics, Volume 6A
© 2007 Elsevier B.V. All rights reserved
DOI: 10.1016/S1573-4412(07)06068-0
Abstract

This chapter covers recent solutions to aggregation problems in three application areas: consumer demand analysis, consumption growth and wealth, and labor participation and wages. Each area involves treatment of heterogeneity and nonlinearity at the individual level. Three types of heterogeneity are highlighted: heterogeneity in individual tastes, heterogeneity in income and wealth risks, and heterogeneity in market participation. Work in each area is illustrated using results from empirical data. The overall aim is to present specific models that connect individual behavior with aggregate statistics, as well as to discuss the principles for constructing such models.

Keywords

aggregation, heterogeneity, consumption growth, consumer demand, wage growth, sample selection

JEL classification: D1, D3, D91, E21, J31, C51
1. Introduction

Models of optimal behavior typically apply at the individual level. Important issues of economic policy typically apply to large groups or entire economies. If different individuals behaved in essentially the same way, then group statistics would mirror that common behavior. However, this is never the case. In virtually every application area, there is evidence of extensive differences in behavior across individuals, or individual heterogeneity. This is true whether the “individual agent” is a single person, a household, a firm or some other decision-making entity that is relevant for the applied question of interest. Individual differences are a fact of life. They create a paramount difference between behavior at the individual level and aggregate statistics for an economy or large group. Resolving that difference involves solving the problem of aggregation.

Some broad properties will transmit between the individual level and aggregate statistics, but typically not enough for quantitative analysis. For instance, consider spending behavior by households when the price of a good increases. If some households decrease their purchases of that good and no household increases its purchases, then aggregate spending on that good must decrease. However, knowing that aggregate spending will decrease is very different from knowing the exact amount it will decrease, or how sensitive aggregate spending is to that price increase. This latter information is required for any analysis of aggregate demand–supply impacts or impacts of policy such as tax changes. To pin down a value to use as the “aggregate price elasticity”, one must come to grips with how individual households differ across the population. In an econometric setting, this requires explicit modeling of household behavior and the differences relevant to spending. Moreover, this example involves the simplest type of change, namely a common price change faced by all households. The issues multiply when the changes themselves vary across households. Consider spending impacts that arise from a policy that injects new income into the household sector. Now it is not even clear how to make sense of a value of an “income elasticity” for the economy. If the policy targets poor households, then one likely gets different impacts than if the policy targeted middle-income households.

The issues raised by aggregation are not new, but rather have been a part of the discussion of empirical work in economics for much of the past century. What is new is the development of econometric models and methods that explicitly deal with aggregation problems. Such models apply equally well to individual data and to aggregate level statistics. The purpose of this chapter is to cover these developments in a few selected application areas. We include discussion of the principles that guide the construction of such econometric models.

In line with the models we cover, we take a practical approach to the definitions of “individual level” and “aggregate statistics”, sidestepping a number of issues about the underpinnings of standard econometric models. For instance, empirical practice in demand modeling is to define a household as the “individual agent” and study total (economy-wide) expenditures on various commodities as the aggregate statistics of in-
Thus, at the micro level, purchase decisions made by individual family members are assumed to be in line with an overall household plan, which sidesteps the issues raised by bargaining within a family. In terms of the macro level, modeling total or “per capita” commodity expenditures entails different issues from modeling an alternative type of aggregate statistic. For instance, different issues would arise if one wanted to study what fraction of households fell below a minimum threshold for food expenditures, or if one wanted to study inequality in living standards from the Gini coefficient of food expenditures.

While practical, this posture is not arbitrary. Much of the development we discuss has been made possible by the enhanced availability of survey data on individual behavior. Standard microeconometric models (and the assumptions that justify them) provide the most natural starting point for building models that account for aggregation. The appropriate choice of aggregate statistics is driven by data availability and the policy question of interest. Economy-wide totals or averages (from national accounts) are the most commonly available statistics for modeling. They are also the most important statistics for most questions of economic policy, such as questions involving prices, interest rates, total savings, market demand and supply, total tax revenues, aggregate wages and unemployment.

Given the application area, an econometric model that accounts for aggregation consists of individual-level equations and equations for aggregate statistics. Ideally, the individual equations will capture all important economic effects and allow for realistic individual heterogeneity. The aggregate equations must be fully consistent with the individual equations, and typically will require assumptions on the distribution of individual heterogeneity. Taken as a whole, these equations constitute a single model that relates to data at all levels – individual cross-section or panel data and aggregate statistics over time. All relevant data sources can be used in estimation, and an estimated model can be applied to any level – individual, proper subgroup or the full economy. There are multiple levels of testable implications of such a model: from the individual model, from the aggregate equations and from the necessary distributional assumptions.

This chapter covers specific models and related work in three application areas: consumer demand analysis, consumption and saving analysis and analysis of wages and labor market participation. A key issue is to identify what kinds of individual differences, or heterogeneity, are relevant for each application area. As an organizing principle, we consider (i) heterogeneity in individual tastes and incomes, (ii) heterogeneity in wealth and income risks faced by individuals and (iii) heterogeneity in market participation.1 There is a generic tension between the degree of individual heterogeneity accounted for and the ease with which one can draw implications for economic aggregates. We point out how different types of heterogeneity are accommodated in the different application areas.

1 This roughly coincides with the categorization of heterogeneity discussed in Browning, Hansen and Heckman (1999).
We are concerned with models that strike a balance between realism (flexibility), adherence to restrictions from economic theory and connections between individual behavior and aggregate statistics. We consider several settings where individual models are intrinsically nonlinear, and for those we must make specific assumptions on the distributions of relevant heterogeneous characteristics. We present results that can be used to explore the impact of heterogeneity in empirical applications, that assume reasonable (and hopefully plausible) parameterizations of both individual equations and distributions of heterogeneity. We do not go into details about estimation; for each application area, we present models with empirically plausible equations for individuals and consistent equations for the relevant economic aggregates. Again, the point is to develop a single framework that has the ability to address empirical issues at the individual (micro) level, the aggregate (macro) level or both.

We begin with our coverage of consumer demand models in Section 2, the area which has seen the most extensive development of solutions to aggregation problems. The difficult issues in consumer demand include clear evidence of nonlinearity in income effects (e.g. Engel’s Law for food) and pervasive evidence of variations in demand with observable characteristics of households. We discuss each of these problems in turn, and use the discussion to cover traditional results as well as “aggregation factors” as a method of empirically studying aggregation bias. We cover recent empirical demand models, and present aggregation factors computed from data on British households. That is, we cover the standard issues faced by aggregating over heterogeneous households in a static decision-making format, and illustrate with application to empirical demand models in current use. We close with a discussion of recent work that studies aggregate demand structure without making specific behavioral assumptions on individual demands.

In Section 3 we discuss models of overall consumption growth and wealth. Here we must consider heterogeneity in tastes, but we focus on the issues that arise from heterogeneity in income shocks, showing how different types of shocks transmit to aggregate consumption. We start with a discussion of quadratic preferences in order to focus on income and wealth, and then generalize to recent empirical models that permit precautionary saving. Because of the log-linear form of these models, we must make explicit distributional assumptions to solve for aggregate equations. We cover the types of heterogeneity found in consumption relationships, as well as various other aspects of our modeling, illustrating with empirical data. We follow this with a brief discussion of modeling liquidity constraints and the impacts on aggregate consumption. We close this section with a discussion of recent progress in general equilibrium modeling of consumption, saving and wealth.

Section 4 covers recent work on labor participation and aggregate wage rates. The main issues here concern how to interpret observed variations in aggregate wages – are they due to changes in wages of individuals or to changes in the population of participating workers? We focus on the issues of heterogeneity in market participation, and develop a paradigm that allows isolation of the participation structure from the wage structure. This involves tracking the impacts of selection on the composition of
the working population, the impacts of weighting individual wage rates by hours in
the construction of aggregate wages and the impact of observed wage heterogeneity.
We show how accounting for these features gives a substantively different picture of the
wage situation in Britain from that suggested by observed aggregate wage patterns. Here
we have a situation where there is substantial heterogeneity and substantial nonlinearity,
and we show how to address these issues and draw conclusions relevant to economic
policy.

Section 5 concludes with some general observations on the status of work on aggre-
gation in economics.

This chapter touches on many of the main ideas that arise in addressing aggregation
problems, but it is by no means a comprehensive survey of all relevant topics or recent
approaches to such problems. For instance, we limit our remarks on the basic nature
of aggregation problems, or how it is senseless to ascribe behavioral interpretations to
estimated relationships among aggregate data without a detailed treatment of the links
between individual and aggregate levels. It is well known that convenient constructs
such as a “representative agent” have, in fact, no general justification – we will not fur-
ther belabor their lack of foundation. See the surveys by Stoker (1993) and Browning,
Hansen and Heckman (1999) for background on these basic problems. It is useful to
mention two related lines of research, that we do not cover. The first is the work on
how economic theory provides few restrictions on market excess demands – see Son-
nenschein (1972) and Schafer and Sonnenschein (1982) among others, and Brown and
Matzkin (1996) for a more recent contribution. The second is the work on collective
decision making within households as pioneered by Chiappori (1988, 1994).

We will also limit our attention to aggregation over individuals, and not discuss the
voluminous literature on aggregation over commodities. This latter literature concerns
the construction of aggregate “goods” from primary commodities, as well as the consist-
tency of multistage budgeting and other simplifications of choice processes. While very
important for empirical work, the issues of commodity aggregation apply within de-
cision processes of individuals and, as such, would take us too far afield of our main
themes. See the survey by Blundell (1988) as well as the book by Blackorby, Pri-
mont and Russell (1978) for background on commodity aggregation and multistage
budgeting. We do not cover the growing literature on hedonic/characteristics models,
which can serve to facilitate commodity aggregation or other simplifications in deci-
sion making. Moreover, we do not cover recent advances that use aggregation to solve
microeconometric estimation problems: see Imbens and Lancaster (1994) for the basic
approach and Berry, Levinsohn and Pakes (2004) for a recent application to estimation
deeds for differentiated products.

Finally, we do not cover in great detail work that is associated with time-series ag-
gregation. That work studies how the time-series properties of aggregate statistics relate
to the time-series processes of associated data series for individuals, such as stationar-
ity and co-integration. To permit such focus, that work relies on strictly linear models
for individual agents, which again turn the discussion away from heterogeneity in in-
dividual reactions and other behavior. We do make reference to time-series properties
of income processes as relevant to our discussion of individual and aggregate consumption, but do not focus on time-series properties in any general way. Interested readers can pursue Granger (1980, 1987, 1990) and the book by Forni and Lippi (1997) for more comprehensive treatment of this literature.2

2. Consumer demand analysis

We begin with a discussion of aggregation and consumer demand analysis. Here the empirical problem is to characterize budget allocation to several categories of commodities. The individual level is that of a household, which is traditional in demand analysis. The economic aggregates to be modeled are average (economy-wide, per household) expenditures on the categories of commodities. We are interested in aggregate demand, or how average category expenditures relate to prices and the distribution of total budgets across the economy.

In a bit more detail, we assume that households have a two-stage planning process, where they set the total budget for the current period using a forward-looking plan, and then allocate that current budget to the categories of nondurable commodities.3 As such, we are not concerned with heterogeneity in the risks faced by households in income and wealth levels—they have already been processed by the household in their choice of total budget (and, possibly, in their stocks of durable goods). We consider commodity categories that are sufficiently broad that household expenditures are nonzero (food categories, clothing categories, etc.), and so we are not concerned with zero responses, or heterogeneity in market participation.

We are concerned with heterogeneity in total budgets and in needs and tastes. It is a well-known empirical fact that category expenditure allocations vary nonlinearly with total budget size (for instance, Engel’s Law with regard to food expenditures). Early applications of exact aggregation demand systems had budget shares in semi-log form (with or without attributes), namely the popular Translog models of Jorgenson, Lau and Stoker (1980, 1982) and Almost Ideal models of Deaton and Muellbauer (1980a, 1980b) respectively. More recent empirical studies have shown the need for further nonlinear terms in certain expenditure share equations. In particular, evidence suggests that quadratic logarithmic income terms are required [see, for example, Atkinson, Gomulka and Stern (1990), Bierens and Pott-Buter (1990), Hausman, Newey and Powell (1994), Hardle, Hildenbrand and Jerison (1991), Lewbel (1991) and Blundell, Pashardes and Weber (1993)]. This nonlinearity means that aggregate demands will be affected by total budget size as well as the degree of inequality in budgets across consumers. It is

---

2 See Stoker (1986c, 1993), Lewbel (1994) and others for examples of clear problems in inferring behavioral reactions from time-series results in the presence of individual heterogeneity.

3 Provided that intertemporal preferences are additive, this accords with a fairly general intertemporal model of expected utility maximization [see Deaton and Muellbauer (1980b), among others].
also well known that category expenditures vary substantially with demographic composition of households, such as how many children are present, or whether the head of household is young or elderly [see Barten (1964), Pollak and Wales (1981), Ray (1983) and Browning (1992)].

Our aim is to understand how behavioral effects for households impinge on price effects and distributional effects on aggregate demands. Understanding these effects is a key ingredient in understanding how the composition of the population affects demand growth over time and relative prices across the different commodity categories.

2.1. Aggregation of consumer demand relationships

Our framework requires accounting for individuals (households), goods and time periods. In each period $t$, individual $i$ chooses demands $q_{ijt}$ (or equivalently expenditures $p_{jt}q_{ijt}$) for $j = 1, \ldots, J$ goods by maximizing preferences subject to an income constraint, where $i = 1, \ldots, n_t$. Prices $p_{jt}$ are assumed to be constant across individuals at any point in time, with $p_t = (p_{1t}, \ldots, p_{Jt})$ summarizing all prices. Individuals have total expenditure budget $m_{it} = \sum_j p_{jt}q_{ijt}$, or income for short, and are described by a vector of household attributes $z_{it}$, such as composition and demographic characteristics. The general form for individual demands is written

$$q_{ijt} = g_{jt}(p_{t}, m_{it}, z_{it}).$$

This model reflects heterogeneity in income $m_{it}$ and individual attributes $z_{it}$. Specific empirical models involve the specification of these elements, including a parametric formula for $g_{jt}$.

Economy-wide average demands and average income are

$$\frac{\sum_i q_{ijt}}{n_t}, \quad j = 1, \ldots, J, \quad \text{and} \quad \frac{\sum_i m_{it}}{n_t}.$$  

We assume that the population of the economy is sufficiently large to ignore sampling error, and represent these averages as the associated population means

$$E_t(q_{ijt}), \quad j = 1, \ldots, J, \quad \text{and} \quad E_t(m_{it}).$$

Our general framework will utilize various other aggregates, such as statistics on the distribution of consumer characteristics $z_{it}$.

---

4 It is common parlance in the demand literature to refer to “total budget expenditure” as “income”, as we do here. In the later section on consumption, we return to using “income” more correctly, as current consumption expenditures plus saving.

5 For most of our discussion, $z_{it}$ can be taken as observable. When we discuss explicit empirical models, we will include unobserved attributes, random disturbances, etc.
2.1.1. Various approaches: Exact aggregation and distributional restrictions

We begin by discussing various approaches to aggregation in general terms. From (1), aggregate demand is given formally as

\[ E_t(q_{ijt}) = \int g_{jt}(p_t, m_{it}, z_{it}) \, dF_t(m_{it}, z_{it}), \quad (4) \]

where \( F_t(m_{it}, z_{it}) \) is the cross-section distribution of income and attributes at time \( t \). At the simplest level, approaches to aggregation seek a straightforward relationship between average demand, average income and average attribute values

\[ E_t(q_{ijt}) = G_{jt}(p_t, E_t(m_{it}), E_t(z_{it})). \quad (5) \]

The exact aggregation approach is based on linearity restrictions on individual preferences/demands \( g_{ijt} \) that allow the relationship \( G_{jt} \) to be derived in a particularly simple way, such that knowledge of \( G_{jt} \) is sufficient to identify (the parameters of) the individual demand model. Take, for example,

\[ g_{jt}(p_t, m_{it}, z_{it}) = b_{0j}(p_t)m_{it} + b_{1j}(p_t)m_{it} \ln m_{it} + b_{2j}(p_t)m_{it}z_{it}, \quad (6) \]

where we suppose \( z_{it} \) is a single variable that has \( z_{it} = 1 \) for an elderly household and \( z_{it} = 0 \) otherwise. Individual demand has a linear term in income and a nonlinear term in income, and the slope of the linear term is different for elderly households. All of these slopes can vary with \( p_t \). Now, aggregate demand is

\[ E_t(q_{ijt}) = b_{0j}(p_t)E_t(m_{it}) + b_{1j}(p_t)E(m_{it} \ln m_{it}) + b_{2j}(p_t)E(m_{it}z_{it}), \quad (7) \]

which depends on average income \( E_t(m_{it}) \) and two other statistics, \( E(m_{it} \ln m_{it}) \) and \( E(m_{it}z_{it}) \). The coefficients are the same in the individual and aggregate models, which is the bridge through which individual preference parameters manifest in aggregate demands (and can be recovered using aggregate data).

In order to judge the impact of aggregation on demand, it is convenient to use aggregation factors.\(^6\) Write aggregate demand as

\[ E_t(q_{ijt}) = b_{0j}(p_t)E_t(m_{it}) + b_{1j}(p_t)\pi_{1t}E(m_{it} \ln m_{it}) + b_{2j}(p_t)\pi_{2t}E(m_{it}z_{it}), \quad (8) \]

where

\[ \pi_{1t} = \frac{E(m_{it} \ln m_{it})}{E(m_{it} \ln m_{it})}, \quad \pi_{2t} = \frac{E(m_{it}z_{it})}{E(m_{it})E(z_{it})}. \quad (9) \]

The factors \( \pi_{1t} \) and \( \pi_{2t} \) show how the coefficients in (7) are adjusted if individual demand is evaluated at average income and average attributes, as in (8). \( \pi_{1t} \) reflects inequality in the income distribution through the entropy term \( E(m_{it} \ln m_{it}) \) and \( \pi_{2t} \)

\(^6\) The use of aggregation factors was first proposed by Blundell, Pashardes and Weber (1993).
reflects the distribution of income of the elderly, as the ratio of the elderly’s share in aggregate income $E(m_{it} z_{it}) / E(m_{it})$ to the percentage of elderly $E(z_{it})$ in the population. Aggregation factors are useful for two reasons. First, if they are stable, then aggregate demand has similar structure to individual demand. Second, their average value indicates how much bias is introduced in estimation using aggregate data alone.\(^7\)

In contrast, the distributional approach considers restrictions on the heterogeneity distribution $F_t(m_{it}, z_{it})$. Suppose the density $dF_t(m_{it}, z_{it})$ is an explicit function of $E_t(m_{it})$, $E(z_{it})$ and other parameters, such as variances and higher-order moments. Then with a general nonlinear specification of individual demands $g_{ijt}$, we could solve (4) directly, expressing aggregate demand $E_t(q_{ijt})$ as a function of those distributional parameters. Here, recovery of individual demand parameters from aggregate demand would be possible with sufficient variation in the distribution $F_t(m_{it}, z_{it})$ over $t$.\(^8\)

While conceptually different from exact aggregation, the distributional approach should not be thought of as a distinct alternative in empirical modeling. With distribution restrictions, formulating a model via direct integration in (4) may be difficult in practice. As such, distributional restrictions are often used together with exact aggregation restrictions, combining simplifying regularities of the income-attribute distribution with linearity restrictions in individual demands.

One example is with mean-scaling, as discussed in Lewbel (1990), where the distribution of income does not change relative shape but just scales up or down. Mean-scaling can arise with a redistribution mechanism where individual budgets are all scaled the same, as in $m_{it} = m_{it-1}(E_t(m_{it}) / E_{t-1}(m_{it-1}))$. This structure allows distributional statistics such as those in (7) to be computed from mean income only.

Another example arises from (distributional) exclusion restrictions. Certain attributes can be excluded from aggregate demand if their distribution conditional on income is stable over time; if

$$
dF_t(m_{it}, z_{it}) = f_z(z_{it}|m_{it}) dF_t^*(m_{it})
$$

where $f_z(z_{it}|m_{it})$ does not vary with $t$, then from (4),

$$
E_t(q_{ijt}) = \int g_{jt}(p_t, m_{it}, z_{it}) f_z(z_{it}|m_{it}) dF_t^*(m_{it})
$$

$$
= \int g_{jt}^*(p_t, m_{it}) dF_t^*(m_{it}).
$$

That is, $z_{it}$ and its distributional statistics are excluded from the equation for aggregate demand. Aggregate demand reflects heterogeneity only through variation in the income

---

\(^7\) For instance, in (8), $b_{1j}(p_{it})$ is the coefficient of $E(m_{it} \ln m_{it})$, whereas $b_{1j}(p_{it})\pi_{it}$ is the coefficient of $E(m_{it})\ln E(m_{it})$. If $\pi_{it}$ is stable, $\pi_{it} = \pi_0$, then $b_{1j}(p_{it})\pi_{it}$ is proportional to $b_{1j}(p_{it})$. In this sense, the structure of aggregate demand matches that of individual demand, but the use of aggregate data alone would estimate the individual coefficient with a proportional bias of $\pi_0$.

\(^8\) Technically, what is necessary for recoverability is completeness of the class of income-attribute distributions; see Stoker (1984a).
distribution – there is not enough variation in the \( z_{it} \) distribution over \( t \) to recover the individual effects from aggregate demand. We discuss various other examples of partial distribution restrictions below.

### 2.1.2. Demand and budget share models

There has been a substantial amount of work on the precise structure of individual preferences and demands consistent with exact aggregation. The most well-known result of this kind is in the extreme case where the aggregate model simply relates average demands \( E_t(q_{ijt}) \) to the vector of relative prices \( p_t \) and average expenditure \( E_t(m_{it}) \). Gorman (1953) showed that this required preferences to be quasi-homothetic, with individual demands linear in \( m_{it} \).

Omitting reference to attributes \( z_{it} \) for now, the general formulation for exact aggregation has demands of the form

\[
q_{ijt} = a_{0j}(p_t) + b_{0j}(p_t) h_0(m_{it}) + \cdots + b_{Mj}(p_t) h_M(m_{it})
\]

with aggregate demands given as

\[
E_t(q_{ijt}) = a_{0j}(p_t) + b_{0j}(p_t) E_t[h_0(m_{it})] + \cdots + b_{Mj}(p_t) E_t[h_M(m_{it})].
\]

As above, provided there is sufficient variation in the statistics \( E_t[h_0(m_{it})], \ldots, E_t[h_M(m_{it})] \), the coefficients \( a_{0j}(p_t), b_{0j}(p_t), \ldots, b_{Mj}(p_t) \), and hence individual demands, can be fully recovered from aggregate data.

Lau (1977, 1982) originally proposed the exact aggregation framework, and demonstrated that demands of the form (12) were not only sufficient but also necessary for exact aggregation, or aggregation without distributional restrictions [c.f. Stoker (1993) and Jorgenson, Lau and Stoker (1982)]. Muellbauer (1975) studied a related problem, and established results for the special case of (12) with only two income terms. These both showed several implications of applying integrability restrictions to (12). If demands are zero at zero total expenditure, then \( a_{0j}(p_t) = 0 \). The budget constraint implies that one can set \( h_0(m_{it}) = m_{it} \), without loss of generality. With homogeneity of degree zero in prices and incomes, one can assert the forms of the remaining income terms, which include the entropy form \( h_1(m_{it}) = m_{it} \ln m_{it} \) and the power form \( h_1(m_{it}) = m_{it}^\theta \). This theory provides the background requirements for specific exact aggregation demand models, such as those we discuss below.

The tradition in empirical demand analysis is to focus on relative allocations, and estimate equations for budget shares. The exact aggregation form (12) is applied to budget shares for this purpose. In particular, if we set \( a_{0j}(p_t) = 0 \) and \( h_0(m_{it}) = m_{it} \)

---

9 Muellbauer (1975) studied the conditions under which aggregate budget shares would depend only on a single representative income value, which turned out to be analogous to the exact aggregation problem with only two expenditure terms.

in (12), then budget shares \( w_{ijt} = \frac{p_{ijt}q_{ijt}}{m_{it}} \) take on a similar linear form. We have
\[
\begin{align*}
    w_{ijt} &= \frac{p_{ijt}q_{ijt}}{m_{it}} = b_{0j}(p_t) + b_{1j}(p_t)h_1(m_{it}) + \cdots + b_{Mj}(p_t)h_M(m_{it}) \\
\end{align*}
\]
(14)
where \( b_{0j}(p_t), \ldots, b_{Mj}(p_t) \) and \( h_1(m_{it}), \ldots, h_M(m_{it}) \) are redefined in the obvious way. If we denote individual expenditure weights as \( \mu_{it} = m_{it}/E_i(m_{it}) \), then aggregate budget shares are
\[
\begin{align*}
    \frac{E_i(p_{ijt}q_{ijt})}{E_i(m_{it})} &= E_i(\mu_{it}w_{ijt}) = b_{0j}(p_t) + b_{1j}(p_t)E_i(\mu_{it}h_1(m_{it})) + \cdots \\
    &+ b_{Mj}(p_t)E_i(\mu_{it}h_M(m_{it})).
\end{align*}
\]
(15)
The same remarks on recoverability apply here: the individual budget share coefficients \( b_{1j}(p_t), \ldots, b_{Mj}(p_t) \) can be identified with aggregate data with sufficient variation in the distributional terms \( E_i(\mu_{it}h_1(m_{it})), \ldots, E_i(\mu_{it}h_M(m_{it})) \) over time. As above, aggregation factors can be used to gauge the difference between aggregate shares and individual shares. We have
\[
\begin{align*}
    \frac{E_i(p_{ijt}q_{ijt})}{E_i(m_{it})} &= E_i(\mu_{it}w_{ijt}) = b_{0j}(p_t) + b_{1j}(p_t)\pi_1E_i(\mu_{it}h_1(m_{it})) + \cdots \\
    &+ b_{Mj}(p_t)\pi_Mh_M(E_i(m_{it}))
\end{align*}
\]
(16)
where by construction
\[
\begin{align*}
    \pi_{kt} = \frac{E_i(\mu_{it}\eta_k(m_{it}))}{h_k(E_i(m_{it}))}, \quad k = 1, \ldots, M,
\end{align*}
\]
(17)
are the aggregation factors. These factors give a compact representation of the distributional influences that cause the aggregate model, and the elasticities derived from it, to differ from the individual model.

The budget share form (14) accommodates exact aggregation through the separation of income and price terms in its additive form. As before, when integrability restrictions are applied to (14), the range of possible model specifications is strongly reduced. A particularly strong result is due to Gorman (1981), who showed that homogeneity and symmetry restrictions imply that the rank of the \( J \times (M + 1) \) matrix of coefficients \([b_{mj}(p_t)]\) can be no greater than 3. Lau (1977), Lewbel (1991) and others have characterized the full range of possible forms for the income functions.

2.1.3. Aggregation in rank 2 and rank 3 models

Early exact aggregation models were of rank 2 (for a given value of attributes \( z_{it} \)). With budget share equations of the form\(^{11}\)
\[
\begin{align*}
    w_{ijt} = b_{0j}(p_t) + b_{1j}(p_t)h_1(m_{it}),
\end{align*}
\]
(18)
\(^{11}\) This is Muellbauer’s (1975) PIGL form.
preferences can be specified that give rise to either the log-form \( h_1(m_{it}) = \ln m_{it} \) or the power form \( h_1(m_{it}) = m_{it}^{\theta} \). Typically the former is adopted and this produces Engel curves that are the same as those that underlie the Almost Ideal model and the Translog model (without attributes).\(^{12}\) In this case, aggregate shares have the form

\[
\frac{E_t(p_{jt}q_{ijt})}{E_t(m_{it})} = E_t(\mu_{it} w_{ijt}) = b_{0j}(p_t) + b_{1j}(p_t)\pi_{1t} \ln E_t(m_{it})
\]

where the relevant aggregation factor is the following entropy measure for the \( m_{it} \) distribution:

\[
\pi_{1t} = \frac{E_t(\mu_{it} \ln m_{it})}{\ln E_t(m_{it})} = \frac{E_t(m_{it} \ln m_{it})}{E_t(m_{it}) \ln E_t(m_{it})},
\]

where we have recalled that \( \mu_{it} = m_{it}/E_t(m_{it}) \). The deviation of \( \pi_{1t} \) from unity describes the degree of bias in recovering (individual) price and income elasticities from aggregate data alone.

Distribution restrictions can be used to facilitate computation of the aggregate statistics as well as studying the aggregation factors. For instance, suppose income is log-normally distributed, with \( \ln m_{it} \) distributed normally with mean \( \mu_{mt} \) and variance \( \sigma^2_{mt} \).

The aggregation factor (20) can easily be seen to be

\[
\pi_{1t} = 1 + \frac{1}{2(\mu_{mt}/\sigma^2_{mt}) + 1}.
\]

To the extent that the log mean and variance are in stable proportion, \( \pi_{1t} \) will be stable. If the log mean is positive, then \( \pi_{1t} > 1 \), indicating positive bias from using \( \ln E_t(m_{it}) \).

Distribution restrictions can also facilitate the more modest goal of a stable relationship between aggregate budget shares and aggregate total expenditure. For instance, suppose that the total expenditure distribution obeys

\[
E_t(m_{it} \ln m_{it}) = c_1 E_t(m_{it}) + c_2 E_t(m_{it}) \ln E_t(m_{it}).
\]

Then aggregate budget shares are

\[
\frac{E_t(p_{jt}q_{ijt})}{E_t(m_{it})} = b_{0j}(p_t) + b_{1j}(p_t)(c_1 + c_2 \ln E_t(m_{it}))
\]

so that a relationship of the form

\[
\frac{E_t(p_{jt}q_{ijt})}{E_t(m_{it})} = \tilde{b}_{0j}(p_t) + \tilde{b}_{1j}(p_t) \ln E_t(m_{it})
\]

would describe aggregate data well.

\(^{12}\) It is worthwhile to note that with the power form, estimation of \( \theta \) with aggregate data would be complicated, because the aggregation statistics would depend in a complicated way on \( \theta \).
Here, integrability properties from individual demands can impart similar restrictions to the aggregate relationship. Lewbel (1991) shows that if individual shares
\[ w_{ijt} = b_0j(p_t) + b_1j(p_t) \ln m_{it} \] (25)
satisfy symmetry, additivity and homogeneity properties, then so will
\[ w_{ijt} = b_0j(p_t) + b_1j(p_t)(\kappa + \ln m_{it}). \] (26)
The analogy of (23) and (26) makes clear that if \( c_2 = 1 \), then the aggregate model will satisfy symmetry, additivity and homogeneity. As such, some partial integrability restrictions may be applicable at the aggregate level.\(^\text{13}\)

As we discuss in Section 2.2 below, rank 2 models of the form (18) fail on empirical grounds. Evidence points to the need for more extensive income effects (for given demographic attributes \( z_{it} \)), such as available from rank 3 exact aggregation specifications. In particular, rank 3 budget share systems that include terms in \((\ln m_{it})^2\) (as well as individual attributes) seem to do a good job of fitting the data, such as the QU AIDS system of Banks, Blundell and Lewbel (1997), described further in Section 2.2 below. In these cases, corresponding to the quadratic term \((\ln m_{it})^2\), there will be an additional aggregation factor to examine,
\[ \pi_{2t} = \frac{E_t(\mu_{it}(\ln m_{it})^2)}{(\ln E_t(m_{it}))^2} = \frac{E_t(m_{it}(\ln m_{it})^2)}{E_t(m_{it})(\ln E_t(m_{it}))^2}. \] (27)
In analogy to (22), one can define partial distributional restrictions so that aggregate shares are well approximated as a quadratic function of \( \ln E_t(m_{it}) \).

2.1.4. Heterogeneous attributes

As we noted in our earlier discussion, the empirical analysis of individual-level data has uncovered substantial demographic effects on demand. Here we reintroduce attributes \( z_{it} \) into the equations, to capture individual heterogeneity not related to income. Since \( z_{it} \) varies across consumers, for exact aggregation, \( z_{it} \) must be incorporated in a similar fashion to total expenditure \( m_{it} \). The budget share form (14) is extended generally to
\[ w_{ijt} = b_0j(p_t) + b_1j(p_t)h_1(m_{it}, z_{it}) + \cdots + b_Mj(p_t)h_M(m_{it}, z_{it}). \] (28)
Restrictions from integrability theory must apply for each value of the characteristics \( z_{it} \). For instance, Gorman’s rank theory implies that the share model can be rewritten

\(^{13}\) It is tempting to consider the case of \( c_1 = 0, c_2 = 1 \), which would imply that the aggregation factor \( \pi_{1t} = 1 \) (and no aggregation bias). However, that case appears impossible, although we do not provide a proof. For instance, if \( m_{it} \) were lognormally distributed, \( c_1 = 0, c_2 = 1 \) would only occur if \( \ln m_{it} \) had zero variance.
with two terms that depend on $m_{it}$, but there is no immediate limit on the number of $h$ terms that depend only on characteristics $z_{it}$.\(^{14}\)

Budget share models that incorporate consumer characteristics in this fashion were first introduced by Jorgenson, Lau and Stoker (1980, 1982). Aggregation factors arise for attribute terms, that necessarily involve interactions between income and attributes. The simplest factors arise for terms that depend only on characteristics, as in $h_j(m_{it}, z_{it}) = z_{it}$, namely

\[
\pi^z_t = \frac{E_t(\mu_{it} z_{it})}{E_t(z_{it})} = \frac{E_t(m_{it} z_{it})}{E_t(m_{it}) E_t(z_{it})}. \tag{29}
\]

This can be seen as the ratio of the income-weighted mean of $z_{it}$ to the unweighted mean of $z_{it}$. If $z_{it}$ is an indicator, say $z_{it} = 1$ for households with children and $z_{it} = 0$ for households without children, then $\pi^z_t$ is the percentage of expenditure accounted for by households with children, $E_t(m_{it} z_{it})/E_t(m_{it})$, divided by the percentage of households with children, $E_t(z_{it})$.

More complicated factors terms arise with expenditure-characteristic effects; for example, if $h_j(m_{it}, z_{it}) = z_{it} \ln m_{it}$ then the relevant aggregation factor is

\[
\pi^z_{1t} = \frac{E_t(\mu_{it} z_{it} \ln m_{it})}{E_t(z_{it}) \ln E_t(m_{it})} = \frac{E_t(m_{it} z_{it} \ln m_{it})}{E_t(m_{it}) E_t(z_{it}) \ln E_t(m_{it})}. \tag{30}
\]

As before, in analogy to (22), one can derive partial distributional restrictions so that aggregate shares are well approximated as a function of $E_t(m_{it})$ and $E_t(z_{it})$.

### 2.2. Empirical evidence and the specification of aggregate demand models

#### 2.2.1. What do individual demands look like?

Demand behavior at the individual household level is nonlinear. As we have mentioned, it is not realistic to assume that demands are linear in total expenditures and relative prices. To illustrate typical shapes of income structure of budget shares, Figures 1 and 2 present estimates of Engel curves of two commodity groups for the demographic group of married couples without children, in the British Family Expenditure Survey (FES).\(^{15}\) Each figure plots the fitted values of a polynomial (quadratic) regression in log total expenditure, together with a nonparametric kernel regression. We see that for food

\(^{14}\) A simple linear transformation will not in general be consistent with consumer optimization. Blundell, Browning and Crawford (2003) show that if budget shares have a form that is additive in functions of $\ln m_{it}$ and demographics, then if (i) Slutsky symmetry holds and (ii) the effects of demographics on budget shares are unrestricted then they have to be linear in $\ln m_{it}$.

\(^{15}\) The FES is a random sample of around 7000 households per year. The commodity groups are nondurable expenditures grouped into: food-in, food-out, electricity, gas, adult clothing, children’s clothing and footwear, household services, personal goods and services, leisure goods, entertainment, leisure services, fares, motoring and gasoline. More precise definitions and descriptive statistics are available on request.
expenditures, an equation that expressed the food share as a linear function of log expenditure would be roughly correct. For alcohol expenditures, the income structure is more complex, requiring quadratic terms in log expenditure. Moreover, as one varies the demographic group, the shapes of the analogous Engel curves are similar, but they vary in level and slope.

The QUAIDS model of Banks, Blundell and Lewbel (1997) seems to be sufficiently flexible to capture these empirical patterns. In the QUAIDS model, expenditure shares have the form

\[ w_{ijt} = \alpha_j + \gamma_j \ln p_t + \beta_j (\ln m_{it} - \ln a(p_t)) + \lambda_j \frac{(\ln m_{it} - \ln a(p_t))^2}{c(p_t)} + u_{ijt} \]

(31)
where $a(p_t)$ and $c(p_t)$ are given as
\[
\ln a(p_t) = \alpha' \ln p_t + \frac{1}{2} \ln p_t' \Gamma \ln p_t,
\]
\[
\ln c(p_t) = \beta' \ln p_t,
\]
with $\alpha = (\alpha_1, \ldots, \alpha_N)'$, $\beta = (\beta_1, \ldots, \beta_N)'$, $\lambda = (\lambda_1, \ldots, \lambda_N)'$ and
\[
\Gamma = \begin{pmatrix} \gamma_1' \\ \vdots \\ \gamma_N' \end{pmatrix}.
\]
This generalizes the (linear) Almost Ideal demand system by allowing nonzero $\lambda_i$ values, with the denominator $c(p_t)$ required to maintain the integrability restrictions. Banks, Blundell and Lewbel (1997) do extensive empirical analysis and establish the importance of the quadratic log expenditure terms for many commodities. Interestingly, they find no evidence of the rejection of integrability restrictions associated with homogeneity or symmetry.

To include demographic attributes, an attractive specification is the ‘shape-invariant’ specification of Blundell, Duncan and Pendakur (1998). Suppose that $g_i^0(\ln m_i)$ denotes a ‘base’ share equation, then a shape-invariant model specifies budget shares as
\[
\hat{w}_{ijt} = g_i^0(\ln m_i - \phi(z_{it}' \theta)) + z_{it} \varphi_j.
\]

The shape-invariant version of the QUAIDS model allows demographic variation in the $\alpha_j$ terms. In Banks, Blundell and Lewbel (1997), the $\alpha_j$, $\beta_j$ and $\lambda_j$ terms in (31) are allowed to vary with many attributes $z_{it}$. Family size, family composition, labor market status, occupation and education are all found to be important attributes for many commodities.17

2.2.2. The implications for aggregate behavior

The stability and interpretation of aggregate relationships can be assessed from examining the appropriate aggregation factors. We can compute the empirical counterparts to the factors by replacing expectations with sample averages. For instance, $\pi_{1t}$ of (20) is estimated as
\[
\hat{\pi}_{1t} = \frac{\sum_j (\hat{\mu}_{ijt} \ln m_{ijt})/n_t}{\ln(\sum_j m_{ijt}/n_t)}.
\]

16 For instance, $\alpha_j + \delta_j' z_{it}$ is used in place of $\alpha_j$, and similar specifications for $\beta_j$ and $\lambda_j$ terms.

17 Various methods can be used to estimate the QUAIDS model, with the iterated moment estimator of Blundell and Robin (2000) particularly straightforward. Banks, Blundell and Lewbel (1997) deal with endogeneity of total expenditures, using various instruments. Finally, we note that Jorgenson and Slesnick (2005) have recently combined a Translog demand model (of rank 3) with an intertemporal allocation model, to model aggregate demand and labor supply in the United States.
and $\pi^*_t$ of (29) is estimated as

$$\hat{\pi}^*_t = \frac{\sum_i \hat{\mu}_{it} z_{it} / n_t}{\sum_i z_{it} / n_t}$$  \hspace{1cm} (33)

where we recall that the weights have the form $\hat{\mu}_{it} = m_{it} / (\sum_i m_{it} / n_i)$. Similarly, quadratic terms in $\ln m_{it}$ will require the analysis of the empirical counterpart to the term (27). Interactions of the $\beta$ and $\lambda$ terms in (31) with demographic attributes necessitate examination of the empirical counterparts of terms of the form (30). We can also study aggregation factors computed over different subgroups of the population, to see how aggregate demand would vary over those subgroups.

Figure 3 presents the estimated $\pi^*_t$ term for the impact of children on household demands. This shows a systematic rise in the share of nondurable expenditures attributable to families with children over the 1980s and 1990s. The aggregate bias associated with using observed percentage of households with children (as opposed to the income distribution across households with and without children) varies from 15% to 25%. The path of $\pi^*_t$ also follows the UK business cycle and the path of aggregate expenditure with downturns in 1981 and 1992.

Figure 4 presents the estimated $\pi_{1t}$ and $\pi_{2t}$ terms relating to the $\ln m_{it}$ and $(\ln m_{it})^2$ expressions in the QUAIDS demand model. It is immediately clear that these also display systematic time-series variation, but in comparison to $\pi^*_t$ above, they increase over the first period of our sample and fall towards the end. The bias in aggregation exhibited for the $(\ln m_{it})^2$ term is more than double that exhibited for the $\ln m_{it}$ term.

Figure 5 presents the aggregation factors for the $\ln m_{it}$ term delineated by certain household types. The baseline $\ln m$ line is the same as that in Figure 4. The other two
Figure 4. Aggregation factors for income structure.

Figure 5. Aggregation factors for income structure by certain household types.

lines correspond to interactions for couples without children and for couples with children. While the time patterns of aggregation factors are similar, they are at different
levels, indicating different levels of bias associated with aggregation over these subgroups.

Finally, it is worthwhile to mention some calculations we carried out on whether distributional restrictions such as (22) are capable of representing the aggregate movements in total expenditure data. Using the time-series of distributional statistics from the FES data, we followed Lewbel (1991) and implemented each of these approximations as a regression. With demographic interaction terms, the aggregate model will only simplify if these conditions also apply to each demographic subgroup. In virtually every case, we found the fit of the appropriate regressions to be quite close (say $R^2$ in the region of 0.99). This gives support to the idea that aggregate demand relationships are reasonably stable empirically. However, the evidence on the $c_j$ terms implies that aggregation factors are substantially different from one, so again, estimates of the price and income elasticities using aggregate data alone will be not be accurate.

2.3. Aggregation of demand without individual structure

We close this section with discussion of a nontraditional approach given in Hildenbrand (1994), which is to study specific aspects of aggregate demand structure without relying on assumptions on the behavior of individual consumers. This work makes heavy use of empirical regularities in the observed distribution of consumer expenditures across the population.

We can understand the nature of this approach from a simple example. Suppose we are interested in whether aggregate demand for good $j$ decreases when price $p_j$ increases (obeying the “Law of Demand”), and we omit reference to other goods and time $t$ for simplicity. Denote the conditional expectation of demand $q_{ij}$ given income $m_i$ and price $p_j$, as $g_j(p_j, m_i)$. Aggregate demand for good $j$ is

$$E(q_{ij}) = G(p_j) = \int g(p_j, m) dF(m)$$

and our interest is in whether $dG/dp_j < 0$. Form this derivative, applying the Slutsky decomposition to $g(p_j, m)$ as

$$\frac{dG}{dp_j} = \int \left[ \frac{dg}{dp_j} \right]_{\text{comp}} - g(p_j, m) \frac{dg}{dm} dF(m)$$

$$= \int \left[ \frac{dg}{dp_j} \right]_{\text{comp}} dF(m) - \int g(p_j, m) \frac{dg}{dm} dF(m)$$

$$= S - A. \quad (34)$$

The price effect on aggregate demand decomposes into the mean compensated price effect $S$ and the mean income effect $A$. If we take $S$ as negative, which is fairly uncontroversial, then we know that $dG/dp_j < 0$ if the income effect $A > 0$. Looking once more at $A$, we can see various ways of ascertaining whether $A > 0$:

$$A = \int g(p_j, m) \frac{dg}{dm} dF(m) = \frac{1}{2} \int \left[ \frac{d[g(p_j, m)]}{dm} \right] dF(m). \quad (35)$$
Without making any structural assumptions on $g(p_j, m)$, one could estimate $A$ with the first expression using nonparametric estimates of $g(\cdot)$ and its derivative. Or, we could examine directly whether the “spread” $g(p_j, m)^2$ is increasing with $m$, and if so, conclude that $A > 0$.

This gives the flavor of this work without doing justice to the details. The main contribution is to link up properties of aggregate demand directly with aspects of the distribution of demands across the population. Hildenbrand (1998) shows that increasing spread is a common phenomenon in data on British households, and it is likely to be valid generally. More broadly, this work has stimulated extensive study of the distribution of household expenditures, with a different perspective from traditional demand modeling. Using nonparametric methods, Hardle, Hildenbrand and Jerison (1991) study aggregate income effects across a wide range of goods, and conclude that the “law of demand” likely holds quite generally. Hildenbrand and Kneip (1993) obtain similar findings on income structure by directly examining the dimensionality of vectors of individual demands.18 See Hildenbrand (1994) for an overview of this work, as well as Hildenbrand (1998) for an examination of variations in the British expenditure distribution within a similar framework.

3. Consumption and wealth

We now turn to a discussion of total consumption expenditures. Here the empirical problem is to characterize consumption expenditures over time periods, including how they relate to income and wealth. The individual level is typically that of a household (or an individual person, depending on data source). The economic aggregate to be modeled is average consumption expenditures over time, and we are interested in how aggregate consumption and saving relate to income and wealth across the economy, as well as to interest rates. This relationship is essential for understanding how interest rates will evolve as the population changes demographically, for instance.

Consumption expenditures are determined through a forward-looking plan, that takes into account the needs of individuals over time, as well as uncertainty in wealth levels. There is substantial evidence of demographic effects and nonlinearities in consumption at the individual level,19 so we will need to consider heterogeneity in tastes as before. Accordingly, aggregate consumption is affected by the structure of households and especially the age distribution, and will also be affected by inequality in the distribution of wealth. We are not concerned here with heterogeneity in market participation per se, as everyone has nonzero consumption expenditures. Later we discuss some issues.

---

18 This is related to the transformation modeling structure of Grandmont (1992). It is clear that the dimensionality of exact aggregation demand systems is given by the number of independent income/attribute terms [c.f. Diewert (1977) and Stoker (1984b)].

19 See Attanasio and Weber (1993a, 1993b) and Attanasio and Browning (1995), among many others.
raised by liquidity constraints, which have much in common with market participation modeling as described in Section 4.

Our primary focus is on heterogeneity with regard to risks in income and wealth levels, and how the forward-planning process is affected by them. We take into account the nature of the income and wealth shocks, as well as the nature of the credit markets that provide insurance against negative shocks.

We consider four different types of shocks, delineated by whether the effects are permanent or transitory, and whether they are aggregate, affecting all consumers, or individual in nature. Aggregate permanent shocks can refer to permanent changes in the productive capability of the economy – such as running out of a key natural resource, or skill-biased technical change – as well as to permanent changes in taxes or other policies that affect saving. Individual permanent shocks include permanent changes in an individual’s ability to earn income, such as chronic bad health and long-term changes in type and status of employment. Aggregate transitory shocks refer to temporary aggregate phenomena, such as exchange rate variation, bad weather and so forth. Individual transitory shocks include temporary job lay-offs, temporary illnesses, etc. Many different situations of uncertainty can be accounted for by combinations of these four different types of shocks.

In terms of risk exposure and markets, there are various scenarios to consider. With complete markets, all risks are insured, and an individual’s consumption path is unaffected by the evolution of the individual’s income over time. When markets are not complete, the extent of available insurance markets becomes important, and determines the degree to which different individual risks are important for aggregate consumption behavior. For example, in the absence of credit market constraints, idiosyncratic risks may be open to self-insurance. But in that case there may be little insurance available for aggregate shocks or even for permanent idiosyncratic shocks. Our discussion takes into account the type of income risks and how risk exposure affects aggregate consumption.

Most of our discussion focuses on individual consumption plans and their implications for aggregate consumption. Beyond this, we can consider the feedback effects on consumption and wealth generated through general equilibrium. For instance, if a certain group of consumers systematically saves more than others, then in equilibrium those consumers will be wealthier, and their saving behavior will be a dominant influence on the evolution of aggregate wealth. The study of this important topic is in its infancy, and has been analyzed primarily with calibrated macroeconomic growth models. We include a discussion of some of this work.

3.1. Consumption growth and income shocks

In our framework, in each period $t$, individual $i$ chooses consumption expenditures $c_{it}$ by maximizing expected utility subject to an asset accumulation constraint. Individual

---

20 See Atkeson and Ogaki (1996) for a model of aggregate expenditure allocation over time and to individual goods based on addilog preferences, assuming that complete markets exist.
has heterogeneous attributes \( z_{it} \) that affect preferences. There is a common, riskless interest rate \( r_t \). We assume separability between consumption and labor supply in each time period, and separability of preferences over time.

We begin with a discussion of aggregation with quadratic preferences. This allows us to focus on the issues of different types of income shocks and insurance, without dealing with nonlinearity. In Section 3.2, we consider more realistic preferences that allow precautionary saving.

When individual within-period utilities are quadratic in current consumption, we have the familiar certainty-equivalent formulation in which there is no precautionary saving. Within-period utilities are given as

\[
U_{it}(c_{it}) = -\frac{1}{2} (a_{it} - c_{it})^2 
\]

for \( c_{it} < a_{it} \). We model individual heterogeneity by connecting \( a_{it} \) to individual attributes as

\[
a_{it} = \alpha + \beta' z_{it}. 
\]

With the discount rate equal to the real interest rate, maximizing the expected sum of discounted utilities gives the following optimal plan for the consumer [Hall (1978)]:

\[
\Delta c_{it} = \Delta a_{it} + \xi_{it} = \beta' \Delta z_{it} + \xi_{it}. 
\]

Defining \( \Omega_{i,t-1} \) as the information set for individual \( i \) in period \( t-1 \), the consumption innovation \( \xi_{it} \) obeys

\[
E[\xi_{it} | \Omega_{i,t-1}] = 0. 
\]

In what follows we will use a time superscript to denote this conditional expectation, namely \( E^t(\cdot) \equiv E[\cdot | \Omega_{i,t-1}] \) to distinguish it from the population average in period \( t \) (which uses a time subscript as in \( E_t(\cdot) \)). Notice, the model (38) is linear in the change in attributes \( \Delta z_{it} \) with constant coefficients \( \beta \), plus the consumption innovation. In other words, this model is in exact aggregation form with regard to the attributes \( z_{it} \) that affect preferences.

### 3.1.1. Idiosyncratic income variation and aggregate shocks

When the only uncertainty arises from real income, the consumption innovation \( \xi_{it} \) can be directly related to the stochastic process for income. We begin by spelling out the income process in a meaningful way. Express income \( y_{it} \) as the sum of transitory and permanent components

\[
y_{it} = y^P_{it} + y^T_{it} 
\]

and assume that the transitory component is serially independent. We assume that the permanent component follows a random walk

\[
y^P_{it} = y^P_{i,t-1} + \eta^P_{it},
\]
where the innovation $\eta_{it}^P$ is serially independent.

Next, decompose these two components into a common aggregate effect and an idiosyncratic effect

\begin{align}
\eta_{it}^P &= \eta_t + \varepsilon_{it}, \\
\gamma_{it}^T &= u_t + v_{it}.
\end{align}

Here $\eta_t$ is the common aggregate permanent shock, $\varepsilon_{it}$ is the permanent shock at the individual level, $u_t$ is the aggregate transitory shock and $v_{it}$ is the individual transitory shock – the four types of income shocks discussed above. This mixture of permanent and transitory shocks has been found to provide a good approximation to the panel data process for log incomes; see Macurdy (1982) and Meghir and Pistaferri (2004). We assume that the individual shocks are normalized to average to zero across the population, namely $E_t(\varepsilon_{it}) = 0$ and $E_t(v_{it}) = 0$.

The stochastic process for individual income then takes the form

\begin{equation}
\Delta y_{it} = \eta_t + \varepsilon_{it} + \Delta u_t + \Delta v_{it}.
\end{equation}

The stochastic process for aggregate income has the form

\begin{equation}
\Delta E_t(y_{it}) = \eta_t + \Delta u_t
\end{equation}

where, again, $E_t$ denotes expectation (associated with averaging) across the population of agents at time $t$.

### 3.1.2. Income shocks and insurance

The first scenario is where individual (and aggregate) shocks are not insurable. Here the optimal consumption innovation $\xi_{it}$ for the individual will adjust fully to permanent income shocks but only adjust to the annuity value of transitory shocks. To see this, again suppose that real interest rates are constant and equal the discount rate. Under quadratic preferences (36), consumption growth can be written [Deaton and Paxson (1994)] as

\begin{equation}
\Delta c_{it} = \beta' \Delta z_{it} + \eta_t + \varepsilon_{it} + \tau_t (\Delta u_t + \Delta v_{it}),
\end{equation}

where $\tau_t$ is the annuitization rate for a transitory shock with planning over a finite horizon.\(^{21}\) Clearly, expected growth is determined by preference attributes as

\begin{equation}
E^{t-1} (\Delta c_{it}) = E(\Delta c_{it} | \Omega_{i,t-1}) = \beta' \Delta z_{it}.
\end{equation}

\(^{21}\) If $L$ is the time horizon, then $\tau_t = r/[1 + r](1 - (1 + r)^{-L-t+1})].$ Clearly $\tau_t \rightarrow 0$ as $r \rightarrow 0$. Note that for a small interest rate, we have $\tau_t \approx 0$, so that the transitory shocks become irrelevant for consumption growth.
Aggregate consumption has the form

\[ \Delta E_t(c_{it}) = \beta' \Delta E_t(z_{it}) + \eta_t + \tau_t \Delta u_t. \]  
\hspace{1cm} (48)

Thus, the aggregate data are described exactly by a representative agent model with quadratic preferences and characteristics \( E_t(z_{it}) \) facing a permanent/transitory income process.\(^{22}\)

For the second scenario, suppose individual shocks can be fully insured, either through informal processes or through credit markets. Now individual consumption growth depends only on aggregate shocks

\[ \Delta c_{it} = \beta' \Delta z_{it} + \eta_t + \tau_t \Delta u_t. \]  
\hspace{1cm} (49)

Consequently, with (48), we will have

\[ \Delta c_{it} = \beta' (\Delta z_{it} - \Delta E_t(z_{it})) + \Delta E_t(c_{it}). \]  
\hspace{1cm} (50)

Thus, consumption growth at the individual level equals aggregate consumption growth plus an adjustment for individual preferences.

Finally, the third scenario is where all shocks (aggregate and individual) are fully insurable. Now individual consumption growth will be the planned changes \( \beta' \Delta z_{it} \) only, and aggregate consumption growth will be the mean of those changes \( \beta' \Delta E_t(z_{it}) \). This is the most complete “representative agent” case, as complete insurance has removed the relevance of all income risks.

### 3.1.3. Incomplete information

It is interesting to note that in our simplest framework, incomplete information can cause aggregate consumption to fail to have random walk structure. In particular, suppose individual shocks are not completely insurable and consumers cannot distinguish between individual and aggregate shocks. To keep it simple, also assume that there are no varying preference attributes \( z_{it} \). Following Pischke (1995), individual \( i \) will view the income process (44) as an MA(1) process:

\[ \Delta y_{it} = \xi_{it} - \theta \xi_{it-1}, \]  
\hspace{1cm} (51)

where the \( \theta \) parameter is a function of the relative variances of the shocks.

Changes in individual consumption are simply

\[ \Delta c_{it} = (1 - \theta) \xi_{it}. \]

\(^{22}\) Aside from the drift term \( \beta' \Delta E_t(z_{it}) \), aggregate consumption is a random walk. In particular, the orthogonality conditions \( E^{t-1}(\eta_t + \tau_t \Delta u_t) = E(\eta_t + \tau_t \Delta u_t|\Omega_{t-1}) = 0 \) hold at the individual level and therefore also hold at the aggregate level.
Note that it is still the case that $E^{t-1}(\Delta c_{it}) = E(\Delta c_{it}|\Omega_{i,t-1}) = 0$. However, from (51) we have that

$$\Delta c_{it} - \theta \Delta c_{it-1} = (1 - \theta) \Delta y_{it}. \tag{52}$$

Replacing $\Delta y_{it}$ by (44) and averaging over consumers we find

$$\Delta E_t(c_{it}) = \theta \Delta E_{t-1}(c_{it-1}) + (1 - \theta)(\eta_t + \Delta u_t) \tag{53}$$

so that aggregate consumption is clearly not a random walk.

### 3.2. Aggregate consumption growth with precautionary saving

With quadratic preferences, consumption growth can be written as linear in individual attributes – in exact aggregation form – and we are able to isolate the impacts of different kinds of income shocks and insurance scenarios. To allow for precautionary saving, we must also account for nonlinearity in the basic consumption process. For this, we now consider the most standard consumption model used in empirical work, that based on Constant Relative Risk Aversion (CRRA) preferences.

#### 3.2.1. Consumption growth with CRRA preferences

We assume that within-period utility is

$$U_{it}(c_{it}) = e^{a_{it}} \left[ \frac{c_{it}^{1 - \frac{1}{s_{it}}}}{1 - \frac{1}{s_{it}}} \right], \tag{54}$$

where $a_{it}$ permits scaling in marginal utility levels (or individual subjective discount rates), and $s_{it}$ is the intertemporal elasticity of substitution, reflecting the willingness of individual $i$ to trade off today’s consumption for future consumption. As before, we will model the heterogeneity in $a_{it}$ and $s_{it}$ via individual attributes $z_{it}$.

We now adopt a multiplicative stochastic income process, with the decomposition expressed in log form as

$$\Delta \ln y_{it} = \eta_t + \varepsilon_{it} + \Delta u_t + \Delta v_{it}. \tag{55}$$

The permanent and transitory error components in the income process are decomposed into aggregate and individual terms, as in (44). As noted before, this income growth specification is closely in accord with the typical panel data models of income or earnings, and it will neatly complement our equations for consumption growth with CRRA preferences. In addition, we assume that the interest rate $r_t$ is small, for simplicity, and is not subject to unanticipated shocks.

With precautionary saving, consumption growth depends on the conditional variances of the uninsurable components of shocks to income. Specifically, with CRRA preferences (54) and log income process (55), we have the following log-linear approximation
for consumption growth:

\[
\Delta \ln c_{it} = \rho r_t + (\beta + \varphi r_t) z_{it} + k_1 \sigma_{it}^{-1} + k_2 \sigma_{At}^{-1} + \kappa_1 \epsilon_{it} + \kappa_2 \eta_t, \tag{56}
\]

where \(\sigma_{it}^{-1}\) is the conditional variance of idiosyncratic risk (conditional on \(t - 1\) information \(\Omega_{it,t-1}\)) and \(\sigma_{At}^{-1}\) is the conditional variance of aggregate risk. The attributes \(z_{it}\) represent the impact of heterogeneity in \(a_{it}\), or individual subjective discount rates, and the intertemporal elasticity of substitution \(s_t = \rho + \beta + \varphi z_{it}\). Typically in empirical applications, \(z_{it}\) will include levels and changes in observable attributes, and unobserved factors may also be appropriate. 

As before,

\[
E(\epsilon_{it} | \Omega_{it,t-1}) = E^{t-1}(\epsilon_{it}) = 0, \tag{57}
\]

\[
E(\eta_t | \Omega_{it,t-1}) = E^{t-1}(\eta_t) = 0. \tag{58}
\]

To sum up, in contrast to the quadratic preference case, the growth equation (56) is nonlinear in consumption, and it includes conditional variance terms which capture the importance of precautionary saving.

A consistent aggregate of the individual model (56) is given by

\[
E_t(\Delta \ln c_{it}) = \rho r_t + (\beta + \varphi r_t) E_t(z_{it}) + k_1 E_t(\sigma_{it}^{-1}) + k_2 \sigma_{At}^{-1} + \kappa_2 \eta_t, \tag{59}
\]

where \(E_t(\Delta \ln c_{it})\) refers to the population mean of the cross-section distribution of \(\Delta \ln c_{it}\) in period \(t\), and so on. The \(t\) subscript again refers to averaging across the population of consumers, and we have normalized \(E_t(\epsilon_{it}) = 0\) as before. Provided \(E_t(\ln c_{it,t-1}) = E_{t-1}(\ln c_{it,t-1})\), Equation (59) gives a model of changes over time in \(E_t(\ln c_{it})\), which is a natural aggregate given the log form of the model (56).

However, \(E_t(\ln c_{it})\) is not the aggregate typically observed nor is it of much policy interest. Of central interest is per-capita consumption \(E_t(c_{it})\) or total consumption \(n_t E_t(c_{it})\). Deriving an equation for the appropriate aggregates involves dealing with the ‘log’ nonlinearity, to which we now turn.

3.2.2. How is consumption distributed?

Since the individual consumption growth equations are nonlinear, we must make distributional assumptions to be able to formulate an equation for aggregate consumption.

\[\text{See Blundell and Stoker (1999) for a precise derivation and discussion of this approximation.}\]

\[\text{See Banks, Blundell and Brugiavini (2001) for a detailed empirical specification of consumption growth in this form.}\]

\[\text{If we evaluate the individual model at aggregate values, we get}\]

\[
\Delta \ln E_t(c_{it}) = \rho r_t + (\beta + \varphi r_t) E_t(z_{it}) + k_2 \sigma_{At}^{-1} + \omega_t.
\]

Here \(\omega_t\) is a ‘catch-all’ term containing the features that induce aggregation bias, that will not satisfy the orthogonality condition \(E^{t-1}(\omega_t) = 0\). It is also worthwhile to note that empirical models of aggregate consumption also typically omit the terms \(E_t(z_{it})\) and \(k_2 \sigma_{At}^{-1}\).
In the following, we will assume lognormality of various elements of the consumption process. Here we point out that this is motivated by an important empirical regularity – namely, individual consumption does appear to be lognormally distributed, at least in developed countries such as the United States and the United Kingdom.

Figure 6 shows the distribution of log-consumption using US consumer expenditure data across the last two decades. Consumption is taken as real expenditure on non-durables and services, and is plotted by five-year bands to achieve a reasonable sample size. Each log-consumption distribution has a striking resemblance to a normal density. In the experience of the authors, this result is often replicated in more disaggregated data by year and various demographic categorizations, such as birth cohort, and also in other countries including in the Family Expenditure Survey data for the UK. Given this regularity, one would certainly start with lognormality assumptions such as those we make below, and any subsequent refinements would need to preserve normality of the marginal distribution of log-consumption.

3.2.3. Insurance and aggregation with precautionary saving

As with our previous discussion, we must consider aggregation under different scenarios of insurance for income risks. We again assume that agents have the same information set, namely $\Omega_{i,t-1} = \Omega_{t-1}$ for all $i, t$.

We begin with the scenario in which there is full insurance for individual risks, or pooling of idiosyncratic risk across individuals. Here insurance and credit markets are sufficiently complete to remove individual risk terms in individual income and consumption streams, so $\epsilon_{it} = 0$ and $\sigma_{it}^{t-1} = 0$ for all $i, t$. The individual model (56)
becomes

\[ \Delta \ln c_{it} = \rho r_t + (\beta + \varphi r_t)z_{it} + k_2 \sigma_{At}^{-1} + \kappa_2 \eta_t \]  
(60)

with \( E_t^{-1}(\eta_t) = 0 \). The mean-log model (59) is now written as

\[ E_t(\ln c_{it}) - E_t(\ln c_{it-1}) = \rho r_t + (\beta + \varphi r_t)E_t(z_{it}) + k_2 \sigma_{At}^{-1} + \kappa_2 \eta_t. \]  
(61)

The relevant aggregate is per-capita consumption \( E_t(c_{it}) \). Per-capita consumption is given by

\[ E_t(c_{it}) = E_t \left[ \exp(\ln c_{it-1} + \rho r_t + (\beta + \varphi r_t)z_{it} + k_2 \sigma_{At}^{-1} + \kappa_2 \eta_t) \right] \]
(62)

with the impact of log-linearity arising in the final term, a weighted average of attribute terms interacted with lagged consumption \( c_{it-1} \).

Of primary interest is aggregate consumption growth, or the log-first-difference in aggregate consumption

\[ \Delta \ln E_t(c_{it}) = \ln \left( \frac{E_t(c_{it})}{E_{t-1}(c_{it-1})} \right). \]
This is expressed as

\[ \Delta \ln E_t(c_{it}) = \rho r_t + k_2 \sigma_{At}^{-1} + \kappa_2 \eta_t \]

\[ + \ln \left( \frac{E_t[c_{it-1} \exp((\beta + \varphi r_t)z_{it})]}{E_t(c_{it-1})} \right) + \ln \left( \frac{E_t(c_{it-1})}{E_{t-1}(c_{it-1})} \right). \]  
(63)

Aggregate consumption growth reflects the interest and risk terms that are common to all consumers, a weighted average of attribute terms, and the log-difference in the average of \( c_{it-1} \) at time \( t \) versus time \( t - 1 \).

Notice first that even if \( z_{it} \) is normally distributed, we cannot conclude that \( \ln c_{it} \) is normal. We also need (as a sufficient condition) that \( \ln c_{it-1} \) is normal at time \( t \) to make such a claim. This would further seem to require normality of \( \ln c_{it-2} \) at \( t - 1 \), and so forth into the distant past. In any case, we cover this situation with the broad assumption:

The distribution of \( c_{it-1} \) is the same in periods \( t - 1 \) and \( t \).  
(64)

That is, the population could grow or shrink, but the distribution of \( c_{it-1} \) is unchanged. Under that assumption, we can drop the last term in (63)

\[ \ln \left( \frac{E_t(c_{it-1})}{E_{t-1}(c_{it-1})} \right) = 0. \]  
(65)

Lagging the individual model (60) gives an equation for \( c_{it-1} \), but there is no natural way to incorporate that structure directly into the equation for aggregate current
consumption $E_t(c_{it})$. Therefore, we further assume

$$
\left( \frac{\ln c_{it-1}}{\theta_i' z_{it}} \right) \sim N \left( \left( \mu_{c_{it-1}}, t \right), \left[ \begin{array}{c}
\theta_i' \Sigma z_{it-1, t} \\
\theta_i' \Sigma z_{z, t}^{\prime, t}
\end{array} \right] \right),
$$

(66)

where we have set $\theta_t = (\beta + \varphi r_t)$. This assumption says that

$$
\ln c_{it-1} + \theta_i' z_{it} \sim \mathcal{N} \left( \mu_{c_{it-1}, t} + \theta_i' E_t(z_{it}), \sigma^2_{c_{it-1}, t} + \theta_i' \Sigma z_{z, t}^{\prime, t} + 2 \theta_i' \Sigma z_{z, t}^{\prime, t} \right)
$$

(67) and

$$
\ln c_{it-1} \sim \mathcal{N} \left( \mu_{c_{it-1}, t}, \sigma^2_{c_{it-1}, t} \right).
$$

(68)

We can now solve for an explicit solution to (63): apply (65), (67) and (68) and rearrange to get

$$
\Delta \ln E_t(c_{it}) = \rho r_t + (\beta + \varphi r_t)' E_t(z_{it}) + k_1 \sigma^2_{At} + k_2 \eta_t + \frac{1}{2} ((\beta + \varphi r_t)' \Sigma z_{z, t}^{\prime, t} (\beta + \varphi r_t) + 2 (\beta + \varphi r_t)' \Sigma z_{z, t}^{\prime, t}).
$$

(69)

This is the aggregate model of interest, expressing growth in per-capita consumption as a function of the mean of $z$, the conditional variance terms from income risk, and the covariances between attributes $z$ and lagged consumption $c_{it-1}$. This shows how individual heterogeneity manifests itself in aggregate consumption through distributional variance terms. These variance terms vary with $r_t$ if the intertemporal elasticity of substitution varies over the population.

Now consider the scenario where some individual risks are uninsurable. This reintroduces terms $\varepsilon_{it}$ and $\sigma_{it}^{t-1}$ in consumption growth at the individual level, and we must be concerned with how those permanent risks are distributed across the population. In particular, we assume in each period that each individual draws idiosyncratic risk from a common conditional distribution, so that $\sigma_{it}^{t-1} = \sigma_{it}^{t-1}$ for all $i$. The individual consumption growth equation (56) now appears as

$$
\Delta \ln c_{it} = \rho r_t + (\beta + \varphi r_t)' z_{it} + k_1 \sigma^2_{At} + k_2 \sigma^2_{At} + \kappa_1 \varepsilon_{it} + \kappa_2 \eta_t.
$$

(70)

The mechanics for aggregation within this formulation are similar to the previous case, including the normalization $E_t(\varepsilon_{it}) = 0$, but we need to deal explicitly with how the permanent individual shocks $\varepsilon_{it}$ covary with $\ln c_{it-1}$. As above, we adopt a stability assumption (64). We then extend (66) to assume that $(\ln c_{it-1}, (\beta + \varphi r_t)' z_{it}, \varepsilon_{it})$ is joint normally distributed. The growth in aggregate average consumption is now given by

$$
\Delta \ln E_t(c_{it}) = \rho r_t + (\beta + \varphi r_t)' E_t(z_{it}) + k_1 \sigma^2_{At} + k_2 \sigma^2_{At} + \kappa_2 \eta_t + \frac{1}{2} (\Lambda_t).
$$

(71)

\footnote{This is because of the potential dependence of $c_{i, t-1}$ on the same factors as $c_{i, t-2}$, and so forth.}
where
\[\Lambda_t = (\beta + \varphi r_t)^2 \Sigma_{\varepsilon,t} (\beta + \varphi r_t) + \kappa^2_t \sigma_{\varepsilon,t}^2 + 2(\beta + \varphi r_t)^2 \Sigma_{\varepsilon^{-1},t} + 2\kappa^1_t \sigma_{I^t}^2 - 2\kappa^1_t \Sigma_{\varepsilon^{-1},t} (\beta + \varphi r_t).\]

While complex, this formulation underlines the importance of the distribution of risk across the population. In contrast to the full information model (69), there is a term \(\sigma_{\varepsilon,t}^2\) in \(\Lambda_t\) that reflects the changing variance in consumption growth. The term \(\sigma_{I^t}^{-1}\) captures how idiosyncratic risk varies, based on \(t - 1\) information.

We have not explicitly considered unanticipated shocks to the interest rate \(r_t\), or heterogeneity in rates across individuals.\(^{27}\) Unanticipated shocks in interest would manifest as a correlation between \(r_t\) and aggregate income shocks, and would need treatment via instruments in estimation. Heterogeneity in rates could, in principle, be accommodated as with heterogeneous attributes. This would be especially complicated if the overall distributional structure were to shift as interest rates increased or decreased.

3.3. Empirical evidence on aggregating the consumption growth relationship

There are two related aspects of empirical research that are relevant for our analysis of aggregation in consumption growth models. The first concerns the evidence on full insurance of individual risks. How good an approximation would such an assumption be? To settle this, we need to examine whether there is evidence of risk pooling across different individuals and different groups in the economy. For example, does an unexpected change in pension rights, specific to one cohort or generation, get smoothed by transfers across generations? Are idiosyncratic health risks to income fully insured? Even though we may be able to cite individual cases where this perfect insurance paradigm clearly fails, is it nonetheless a reasonable approximation when studying the time-series of aggregate consumption?

The second aspect of empirical evidence concerns the factors in the aggregate model (71) that are typically omitted in studies of aggregate consumption. From the point of view of estimating the intertemporal elasticity parameter \(\rho\), how important are these aggregation factors? How well do they correlate with typically chosen instruments and how likely are they to contaminate tests of excess sensitivity performed with aggregate data?

3.3.1. Evidence on full insurance and risk pooling across consumers

If the full insurance paradigm is a good approximation to reality, then aggregation is considerably simplified and aggregate relationships satisfying the standard optimality conditions can be derived with various conditions on individual preferences. There is a reasonably large and expanding empirical literature on the validity of the full insurance

\(^{27}\) Zeldes (1989b) points out how differing marginal tax rates can cause interest \(r_t\) to vary across consumers.
scenario, as well as complete markets scenario. This work is well reviewed in Attanasio
(1999) and Browning, Hansen and Heckman (1999). Here we present evidence directly
related to our discussion of consumption growth above. Two rather effective ways of
analyzing failures of the full insurance paradigm fit neatly with our discussion.

One approach to evaluating the full insurance hypothesis is to look directly for ev-

dence that unexpected shocks in income across different groups in the economy lead
to differences in consumption patterns (as consistent with (56), which assumes no in-
surance). This is not a trivial empirical exercise. First, such income shocks have to be
identified and measured. Second, there has to be a convincing argument that they would
not be correlated with unobservable variables entering marginal utility, or observables
such as labor supply (in a nonseparable framework).

Building on the earlier work by Cochrane (1991), Mace (1991), Hayashi, Altonji
and Kotlikoff (1996) and Townsend (1994), the study by Attanasio and Davis (1996)
presents rigorous and convincing evidence against the full insurance hypothesis us-
ing this approach. Low-frequency changes in wages across different education and
date-of-birth cohorts are shown to be correlated positively with systematic differences
in consumption growth. More recently, Blundell, Pistaferri and Preston (2003) use a
combination of the Panel Survey of Income Dynamics (PSID) and the Consumers Ex-
penditure Survey (CES) to investigate insurance of permanent and transitory income
shocks at the individual level. They find almost complete insurance to transitory shocks
except among lower-income households. They find some insurance to permanent shocks
particularly among the younger and higher educated. But they strongly reject the com-
plete insurance model.

The second approach to evaluating full insurance is to assume risk-averse preferences
and to model the evolution of idiosyncratic risk terms. In terms of the model (56), this
approach examines the relevance of individual risk terms (e.g. $\sigma_{it}^{-1}$) once aggregate
risk ($\sigma_{At}^{-1}$) has been allowed for. This is addressed by looking across groups where the
conditional variance of wealth shocks is likely to differ over time and to see whether
this is reflected in differences in consumption growth. Following earlier work by Dynan
(1993), Blundell and Stoker (1999), Caballero (1990) and Skinner (1988), the study
by Banks, Blundell and Brugiavini (2001) presents evidence that differential variances
of income shocks across date-of-birth cohorts do induce important differences in con-
sumption growth paths.

3.3.2. Aggregation factors and consumption growth

There are two issues. First, if one estimates a model with aggregate data alone, is there
likely to be bias in the estimated parameters of interest? Second, will the omission of
aggregation bias terms result in spurious inference concerning the presence of excess
sensitivity of consumption to transitory income shocks?

With regard to bias, we consider the elasticity of intertemporal substitution $\rho$, which
is normally a focus of studies of aggregate consumption. In Figure 7 we plot the aggre-
Figure 7. Aggregation factor for consumption growth.

The study of excess sensitivity involves the use of lagged information as instrumental variables in the estimation of the consumption growth relationship. Omitting aggregation bias terms can invalidate the instruments typically used. For the consumption data used above, we computed the correlation of the aggregation factor with two typically used instrumental variables in consumption growth equations – lagged real interest rates

\[ \Delta \ln E_t(c_{it}) - \Delta E_t(\ln c_{it}) \]  

for the sample of married couples from the British FES, used to construct the aggregation factors for demand of Section 2. The figure shows a systematic procyclical variation. We found the correlation coefficient between the real interest series and this factor to be significant. This indicates that there will exist an important aggregation bias in the estimated intertemporal substitution parameter from aggregate consumption data (with a log-linear growth model). This is confirmed in the study by Attanasio and Weber (1993a, 1993b), where aggregate data was constructed from micro survey information. They find an elasticity estimate for aggregate data of around 0.35, and the corresponding micro-level estimates were twice this size.

\(28\) Attanasio and Weber (1993a, 1993b) also note a strong impact of omitting the cross-section variance of consumption growth.
and lagged aggregate consumption. The estimated correlation coefficient between these series and the omitted bias term was found to be strongly significant.\footnote{Detailed regression results available on request.}

Together these results suggest that aggregation problems are likely to lead to serious bias in estimated intertemporal substitution parameters and also to exaggerate the presence of excess sensitivity in consumption growth regressions on aggregate data. Attanasio and Browning (1995) investigate this excess sensitivity issue in more detail and find that excess sensitivity still exists at the micro-data level but disappears once controls for age, labour supply variables and demographics are introduced in a flexible way. Moreover, these variables explain why excess sensitivity appears to vary systematically over the cycle.

It is an important finding that evidence of excess sensitivity vanishes once we move to individual data and include observable variables that are likely to impact preferences for the allocation of consumption over time. It has important consequences for our understanding of liquidity constraints and for partial insurance. It has implications for understanding the path of consumption growth over the cycle. It also has implications for the retirement-savings puzzle, or how consumption drops much more at retirement than is predicted by standard consumption growth equations. Banks, Blundell and Tanner (1998) find that once demographics and labor supply variables are allowed to affect the marginal utility of consumption, nearly two thirds of the retirement-savings puzzle disappears.

### 3.4. Consumption and liquidity constraints

Our previous discussion has focused on heterogeneity in wealth and income risk as it impinges on consumption. We now turn to a discussion of liquidity constraints on consumption, which generate a different kind of aggregation structure. The evidence for liquidity constraints is relatively limited. Most studies of consumption smoothing at the individual level find it difficult to reject the standard model once adequate care is taken in allowing for demographic and labor market interactions; see Attanasio and Weber (1993a, 1993b) and Blundell, Browning and Meghir (1994), for example. Much of the excess sensitivity found in aggregate studies can be attributed to aggregation bias as documented in Attanasio and Weber (1993a, 1993b), Goodfriend (1992) and Pischke (1995). However, there is some evidence that does point to the possibility that a fraction of consumers could be liquidity constrained at particular points in the life-cycle and business cycle. At the micro level some evidence can be found in the studies by Hayashi (1987), Zeldes (1989a), Jappelli (1990), Jappelli and Pagano (1994), Meghir and Weber (1996) and Alessie, Devereux and Weber (1997). As mentioned earlier, the Blundell, Pistaferri and Preston (2003) study shows that the consumption of low-income households in the PSID does react to transitory shocks to income, which suggests that such households do not have access to credit markets to smooth such shocks.
For aggregation, liquidity constraints introduce regime structure into the population. Namely, liquidity-constrained consumers constitute one regime, unconstrained consumers constitute another regime, and aggregate consumption will depend upon the relative distribution across regimes. This structure is particularly relevant for the reaction of consumption growth to increases in current income, since constrained consumers will show a stronger reaction than unconstrained consumers. In this section we discuss these basic issues, and indicate how a model of aggregates can be constructed. Blundell and Stoker (2003) work out the details for aggregate consumption models of this type.

There is some subtlety in considering what population groups are likely to be liquidity constrained. Poor households with a reasonably stable but low expected stream of income may have little reason to borrow. More likely to be constrained are young consumers, who have much human capital but little financial wealth—college students or perhaps poor parents of able children. Such individuals may want to borrow against their future earned incomes but cannot, in part, because their eventual income is higher than others', and the growth of their income with experience is higher. Clearly such consumers will react more than others to shocks in current income and wealth.

We start with the basic consumption model discussed earlier, with permanent and transitory shocks to income. As in (55), the change in current income for consumer \( i \) at time \( t \) is

\[
\Delta \ln y_{it} = \eta_t + \varepsilon_{it} + \Delta u_t + \Delta v_{it},
\]

(73)

where \( \eta_t + \varepsilon_{it} \) is the permanent component and \( \Delta u_t + \Delta v_{it} \) is the transitory component.

To keep things simple, we assume that permanent income shocks are not insurable, with log-consumption given as

\[
\Delta \ln c_{it} = \rho r_t + (\beta + \varphi r_t)z_{it} + \eta_t + \varepsilon_{it},
\]

(74)

where we assume the precautionary risk terms \( \sigma_{\text{it}}^{\tau - 1}, \sigma_{\text{At}}^{\tau - 1} \) are included with the \( z_{it} \) effects. Note that (74) gives the consumption growth plan \( (\rho r_t + (\beta + \varphi r_t)z_{it}) \) as well as how consumption reacts to permanent shocks in income (here \( \eta_t + \varepsilon_{it} \)).

Liquidity constraints affect the ability of consumers to finance their desired consumption growth path. We follow an approach similar to Zeldes (1989a), where the incidence of liquidity constraints depends on the degree of consumption growth the consumer is trying to finance and the existing stock of assets. In particular, liquidity constraints enter the growth plan only if they are binding in planning period \( t - 1 \), and then the best response will always be to increase consumption growth so as to “jump” back up to the optimal path. If this response is further frustrated by a binding constraint in period \( t \), consumption will simply grow by the amount of resources available.

This response structure is captured by additional terms in Equation (74). Let \( I_{it} \) denote the indicator

\[
I_{it} = 1[\text{consumer } i \text{ is constrained in period } t - 1]
\]

(75)
and suppose that a consumer who is constrained in period \( t - 1 \) needs to increase consumption growth by \( m_{it} \) to return to the optimal growth plan. Then, consumption growth for unconstrained consumers is

\[
\Delta \ln c_{it} = \rho r_t + (\beta + \varphi r_t)z_{it} + I_{it}m_{it} + \eta_t + \epsilon_{it}.
\]  

(76)

We now model the constraints, as well as consumption growth for constrained consumers. With growth in income of \( \Delta \ln y_{it} \), consumer \( i \) needs to finance a growth rate of

\[
\rho r_t + (\beta + \varphi r_t)z_{it} + I_{it-1}m_{it} + \eta_t + \epsilon_{it} - \Delta \ln y_{it}
\]

\[
= \rho r_t + (\beta + \varphi r_t)z_{it} + I_{it}m_{it} - \Delta u_t - \Delta v_{it}
\]

for consumption at time \( t \) to be on the growth plan. To model liquidity constraints at time \( t \), suppose that consumer \( i \) faces a borrowing constraint that is associated with a maximum rate of increase of consumption of

\[
\gamma + \delta A_{it} + \zeta_{it},
\]

where \( A_{it} \) is (say) accumulated financial wealth. Consumer \( i \) is liquidity constrained in period \( t \), or cannot maintain the consumption growth plan, if

\[
\rho r_t + (\beta + \varphi r_t)z_{it} + I_{it-1}m_{it} - \Delta u_t - \Delta v_{it} > \gamma + \delta A_{it} + \zeta_{it}
\]  

(77)

which we indicate by \( I_{it} = 1 \), as above. In this case we assume that consumption growth is as large as possible, namely

\[
\Delta \ln c_{it} = \Delta \ln y_{it} + \gamma + \delta A_{it} + \zeta_{it}.
\]  

(78)

In terms of permanent and transitory terms of income growth, (78) may be rewritten as

\[
\Delta \ln c_{it} = \eta_t + \epsilon_{it} + \Delta u_t + \Delta v_{it} + \gamma + \delta A_{it} + \zeta_{it}.
\]  

(79)

This is consumption growth for constrained consumers. The constraints have an impact; as consumption growth clearly depends on transitory income shocks and wealth levels.

Aggregate consumption growth will clearly depend on the proportion of consumers who are constrained and the proportion that are not. Consumers who were constrained last period will have a boost in their consumption growth to return to the optimal path. This regime-switching structure is nonlinear in character. Therefore, to model aggregate consumption growth, we would need to specify distributional structure for all the elements that are heterogeneous across the population. We then aggregate over the population of unconstrained individuals with consumption growth (76) and the population of constrained individuals with consumption growth (79). Using log-normality assumptions, we carry out this development in Blundell and Stoker (2003). It is clear how aggregate consumption is affected by transitory income shocks, as well as the distribution of wealth.

---

30 Various approaches have been applied to account for the jump term \( m_{it} \) in studies of micro-level data. See Zeldes (1989a), Jappelli, Pischke and Souleles (1998), Garcia, Lusardi and Ng (1997), Alessie, Melenberg and Weber (1988), Alessie, Devereux and Weber (1997) and Attanasio and Weber (1993a, 1993b).
3.5. Equilibrium effects

As we mentioned at the start, one use of aggregate consumption equations is to study and understand the evolution of aggregate consumption and saving by themselves. Another important use is in studying equilibrium price and interest rate paths over time. This is an exercise in general equilibrium analysis, and every feature that we have discussed above is relevant — consumer heterogeneity, heterogeneity in income and wealth risks, liquidity constraints, and the distribution of wealth. Further complicating this effort is the dynamic feedback that occurs wherein the level and distribution of wealth evolve as a result of the level and distribution of saving. These difficulties make it very hard to obtain analytical results on equilibrium. Nevertheless, it is extremely important to understand the nature of equilibrium here, including implications on prices and interest rates. We now discuss some recent progress that has been made using calibrated stochastic growth models. A leading example of this effort is provided by Krusell and Smith (1998), although the approach dates from at least Aiyagari (1994a, 1994b) and Heaton and Lucas (1996).

The Krusell–Smith setup has the following features. Consumers are infinitely-lived, with identical (within-period) CRRA preferences, but they are heterogeneous with regard to discount rates. Each consumer has a probability of being unemployed each period, providing transitory, idiosyncratic income shocks. Production arises from a constant returns-to-scale technology in labor and capital, and productivity shocks provide transitory aggregate shocks. Consumers can insure by investing in capital only, so that insurance markets are incomplete, and consumers’ capital holdings cannot be negative (liquidity constraint). This setup is rich but in many ways is very simple. Nevertheless, in principle, in order to predict future prices, each consumer must keep track of the evolution of the entire distribution of wealth holdings.

Krusell and Smith’s simulations show a rather remarkable simplification to this forecasting problem. For computing equilibrium and for consumer planning, it is only necessary for consumers to keep track of two things, the mean of the wealth distribution and the aggregate productivity shock. Thus there is an informational economy afforded in a similar fashion to a formal aggregation result: once mean wealth is known, the information contained in the distribution of wealth does not appear to improve forecasting very much. This is true even with heterogeneity of many types, including individual and aggregate income shocks (albeit transitory).

The reason for this is clear once the nature of equilibrium is examined. Most consumers, especially those with lowest discount rates, save enough to insure their risk to the point where their propensity to save out of wealth is essentially constant and unaffected by current income or output. Those consumers also account for a large fraction of the wealth. Therefore, saving is essentially a linear function in wealth, and only the mean of wealth matters to how much aggregate saving is done each period. The same is not true of aggregate consumption. There are many low-wealth consumers who become unemployed and encounter liquidity constraints. Their consumption is much more sensitive to current output than that of wealthier consumers. In essence what is happening
here is that the dynamics of the saving process concentrates wealth in the hands of a
group that behaves in a homogeneous way, with a constant marginal propensity to save.
This (endogenous) simplification allows planning to occur on the basis of mean wealth
only.

It is certainly not clear how applicable this finding is beyond the context of this study.
This is a computational finding that depends heavily on the specifics of this particular
setup. Nonetheless, this form of feedback has some appeal as an explanation of the
smooth evolution of wealth distribution, as well as why forecasting equations that fit
well are so often much simpler than one would expect from the process that underlies
the data. The rich are different (and in this model, the difference makes them rich), but
what is important for forecasting is how similar the rich are to one another. With equal
saving propensities, it does not matter which group of rich people holds the most wealth.

The study of equilibrium effects and aggregation is in its infancy. Of further note are
recent attempts to model differences in micro and macro labor supply elasticities. This
includes Chang and Kim (2006) and Rogerson and Wallenius (2007), who incorporate
individual decisions at the extensive margin, such as labor participation. In the next
section, we discuss labor participation and selection from a partial equilibrium perspec-
tive. In any event, we expect the study of equilibrium effects to generate many valuable
insights.

4. Wages and labor participation

Our final topic area is the analysis of wages and labor participation. Here the empirical
problem is to understand the determinants of wages separately from the determinants
of participation. The individual level is that of an individual worker. The economic ag-
gregates to be modeled are aggregate wages and the aggregate participation rate, or one
minus the unemployment rate. These statistics are central indicators for macroeconomic
policy and for the measurement of economic well-being.

Our analysis is based on a familiar paradigm from labor supply. Potential wages are
determined through human capital, and labor participation is determined by comparing
potential wages to a reservation wage level. Empirically, there is substantial heterogene-
ity in the determinants of wages, and substantial heterogeneity in the factors determining
labor participation, and both processes are nonlinear. In particular, it is typical to specify
wage equations for individuals in log form, and there is much evidence of age and co-
hort effects in wages and employment. As with demand and consumption, we will need
to be concerned with heterogeneity in individual attributes. To keep things as simple as
possible, we do not consider forward-looking aspects of employment choice, and so are
not concerned with heterogeneity in income and wealth risks.

31 Carroll (2000) makes a similar argument, with emphasis on the role of precautionary saving. Krusell and
Smith (2007) survey recent work, arguing that their original findings are robust to many variations in their
framework.
Our primary focus is on heterogeneity in market participation. Aggregate wages depend on the rate of participation, and the important issues involve separation of the wage process from the participation decision. To put it very simply, suppose aggregate wages are increasing through time. Is this because typical wages for workers are increasing? Or, is it because low-wage individuals are becoming unemployed? Do the sources of aggregate wage growth vary other the business cycle? The aggregation problem must be addressed to answer these questions.

We now turn to our basic model of wages that permits us to highlight these effects. We then show the size of these effects for aggregate wages in the UK, a country where there have been large and systematic changes in the composition of the workforce and in hours of work. A more extensive version of this model and the application is given in Blundell, Reed and Stoker (2003). They also summarize derivations of all aggregate equations given below.

4.1. Individual wages and participation

We begin with a model of individual wages in the style of Roy (1951), where wages are based on human capital or skill levels, and any two workers with the same human capital level are paid the same wage. Our framework is consistent with the proportionality hypothesis of Heckman and Sedlacek (1990), where there is no comparative advantage, no sectoral differences in wages for workers with the same human capital level, and the return to human capital is not a function of human capital endowments.

We assume that each worker possesses a human capital (skill) level of $H_i$. Suppose human capital is nondifferentiated, in that it commands a single price $r_t$ in each time period $t$. The wage paid to worker $i$ at time $t$ is

$$ w_{it} = r_t H_i. \quad (80) $$

Human capital $H_i$ is distributed across the population with mean

$$ E_t(\ln H_i) = \delta_{js} $$

where $\delta_{js}$ is a level that varies with cohort $j$ to which $i$ belongs and education level $s$ of worker $i$. In other words, the log-wage equation has the additive form

$$ \ln w_{it} = \ln r_t + \delta_{js} + \varepsilon_{it} \quad (81) $$

where $\varepsilon_{it}$ has mean 0. We will connect $\delta_{js}$ to observable attributes below.

To model participation, we assume that reservation wages $w^*_it$ are lognormal:

$$ \ln w^*_it = \alpha \ln B_{it} + \eta_{js} + \zeta_{it}, \quad (82) $$

32 Heckman and Sedlacek (1985) provide an important generalization of this framework to multiple sectors. See also Heckman and Honore (1990).

33 Clearly, there is an indeterminacy in the scaling of $r_t$ and $H_i$. Therefore, to study $r_t$, we will normalize $r_t$ for some year $t = 0$ (say to $r_0 = 1$). We could equivalently set one of the $\delta$s to zero.
where $\zeta_{it}$ has mean 0 and where $B_{it}$ is an exogenous income (welfare benefit) level that varies with individual characteristics and time. Participation occurs if $w_{it} \geq w^*_it$, or with

$$\ln r_t - \alpha \ln B_{it} + \delta_js - \eta_js - \xi_{it} \geq 0.$$  (83)

We represent the participation decision by the indicator $I_{it} = 1[w_{it} \geq w^*_it]$.

For aggregation over hours of work, it is useful to make one of two assumptions. One is to assume that the distribution of hours is fixed over time. The other is to assume that desired hours $h_{it}$ are chosen by utility maximization, where reservation wages are defined as $h_{it}(w^*) = h_0$ and $h_0$ is the minimum number of hours available for full-time work. We assume $h_{it}(w)$ is normal for each $w$, and approximate desired hours by

$$h_{it} = h_0 + \gamma \cdot (\ln w_{it} - \ln w^*_it).$$  (84)

This is our base-level specification. It is simple to extend this model to allow differentiated human capital, or differential cohort effects due to different labor market experience, which permits a wide range of education/cohort/time effects to be included [c.f. Blundell, Reed and Stoker (2003)]. Because our examples involve log-linear equations and participation (or selection), we summarize the basic framework as

$$\ln w_{it} = \beta_0 + \beta'x_{it} + \epsilon_{it},$$

$$I_{it} = 1[\alpha_0 + \alpha'z_{it} + v_{it} \geq 0],$$

$$h_{it} = h_0 + \gamma \cdot (\alpha_0 + \alpha'z_{it} + v_{it}).$$  (85)

Here, $x_{it}$ denotes education, demographic (cohort, etc.) and time effects, $z_{it}$ includes out-of-work benefit variables, and $I_{it} = 1$ denotes participation. It is clear that the scale of $\gamma$ is not identified separately from the participation index $\alpha_0 + \alpha'z_{it} + v_{it}$; however, we retain $\gamma$ to distinguish between the fixed hours case $\gamma = 0$ and the variable hours case $\gamma \neq 0$.

Our notation distinguishes two types of individual heterogeneity in (85). The variables $x_{it}$ and $z_{it}$ are observable at the individual level, while $\epsilon_{it}$ and $v_{it}$ are unobservable. Analysis of data on wages and participation at the individual level requires assumptions on the distribution of those unobservable elements, a process familiar from the literature on labor supply and selection bias. We now review some standard selection formulae here for later comparison with the aggregate formulations. Start with the assumption that the unobserved elements are normally distributed

$$\begin{pmatrix} \epsilon_{it} \\ v_{it} \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_\epsilon & \sigma_{\epsilon v} \\ \sigma_{\epsilon v} & \sigma^2_v \end{pmatrix} \right).$$  (86)

This allows us to apply some well-known selection formulae (given in virtually every textbook of econometrics). The micro participation regression, or the proportion of par-
participants given \( x_{it} \) and \( z_{it} \), is a probit model:

\[
E_t[I|x_{it}, z_{it}] = \Phi \left( \frac{\alpha_0 + \alpha'z_{it}}{\sigma_z^2} \right). \tag{87}
\]

The micro log-wage regression for participants is

\[
E_t[\ln w_{it}|I_{it} = 1, x_{it}, z_{it}] = \beta_0 + \beta'x_{it} + \frac{\sigma_y}{\sigma_y^2} \lambda \left( \frac{\alpha_0 + \alpha'z_{it}}{\sigma_z^2} \right) \tag{88}
\]

reflecting the typical (Heckman-style) selection term, which adjusts the log-wage equation to the group of participating workers.34

4.2. Aggregate wages and employment

The aggregate of interest is average hourly earnings, where aggregation occurs over all workers, namely

\[
\bar{w}_t = \frac{\sum_{i \in (I = 1)} h_{it} w_{it}}{\sum_{i \in (I = 1)} h_{it}} = \sum_{i \in (I = 1)} \mu_{it} w_{it}, \tag{89}
\]

where \( i \in (I = 1) \) denotes a participant (worker), \( h_{it} w_{it} \) is the earnings of individual \( i \) in period \( t \), and \( \mu_{it} \) are the hours-weights

\[
\mu_{it} = \frac{h_{it}}{\sum_{i \in (I = 1)} h_{it}}.
\]

Modelling the aggregate wage (89) requires dealing with log-nonlinearity of the basic wage equation, dealing with participation and dealing with the hours-weighting. All of these features require that distributional assumptions be made for (observable) individual heterogeneity. In particular, we make the following normality assumption for \( x_{it} \) and \( z_{it} \):

\[
\left( \begin{array}{c} \beta_0 + \beta'x_{it} \\ \alpha_0 + \alpha'z_{it} \end{array} \right) \sim \mathcal{N} \left( \left( \begin{array}{c} \beta_0 + \beta' \mathbb{E}(x_{it}) \\ \alpha_0 + \alpha' \mathbb{E}(z_{it}) \end{array} \right), \left( \begin{array}{cc} \beta' \Sigma_{xx} \beta & \alpha' \Sigma_{xz} \beta \\ \alpha' \Sigma_{xz} \alpha & \alpha' \Sigma_{zz} \alpha \end{array} \right) \right) \tag{90}
\]

or that the indices determining log-wage and employment are joint normally distributed.35

We now discuss some aggregate analog of the micro regression equations, and then our final equation for the aggregate wage. The aggregate participation (employment)

---

34 Here \( \Phi(\cdot) \) is the normal cumulative distribution function, and \( \lambda(\cdot) = \phi(\cdot)/\Phi(\cdot) \), where \( \phi(\cdot) \) is the normal density function.

35 Assuming that the linear indices are normal is much weaker than assuming that \( x_{it} \) and \( z_{it} \) are themselves joint multivariate normal. Such a strong structure would eliminate many important regressors, such as qualitative variables.
rate is
\[ E_t[I] = \Phi \left[ \alpha_0 + \alpha' E(z_{it}) / \sqrt{\alpha' \Sigma_{zz} \alpha + \sigma^2_\nu} \right], \]
(91)
using a formula originally due to McFadden and Reid (1975). Aggregate participation has the same form as the micro participation regression (87) with \( z_{it} \) replaced by \( E(z_{it}) \) and the spread parameter \( \sigma^2_\nu \) replaced by the larger value \( \sqrt{\alpha' \Sigma_{zz} \alpha + \sigma^2_\nu} \), reflecting the influence of heterogeneity in the individual attributes that affect the participation decision. The mean of log-wages for participating (employed) workers is
\[ E_t[\ln w_{it} | I_{it} = 1] = \beta_0 + \beta' E_t(x_{it} | I = 1) \]
(92)
using a formula originally derived by MaCurdy (1987). This matches the micro log-wage regression (88) with \( x_{it} \) replaced by \( E_t(x_{it} | I = 1) \), \( z_{it} \) replaced by \( E_t(z_{it}) \) and the spread parameter changed from \( \sigma^2_\nu \) to \( \sqrt{\alpha' \Sigma_{zz} \alpha + \sigma^2_\nu} \). This is an interesting result, but does not deliver an equation for the aggregate wage \( \bar{w}_t \).

Blundell, Reed and Stoker (2003) derive such an equation. The aggregate wage is given as
\[ \ln \bar{w}_t = \ln \frac{E_t[h_{it} w_{it} | I_{it} = 1]}{E_t[h_{it} w_{it} | I_{it} = 1]} = \beta_0 + \beta' E_t(x_{it} | I = 1) + [\Omega_t + \Psi_t + \Lambda_t], \]
(93)
where the aggregation bias is comprised of a spread term
\[ \Omega_t = \frac{1}{2} \left[ \beta' \Sigma_{xx} \beta + \sigma^2_\epsilon \right], \]
(94)
plus two terms \( \Psi_t \) and \( \Lambda_t \), which represent separate sources of bias but have very complicated expressions.36

What these terms represent can be seen most easily by the following construction. Begin with the individual wage equation evaluated at mean attributes, \( \beta_0 + \beta' E_t(x_{it}) = E_t(\ln w_{it}) \), or overall mean log-wage. Adding \( \Omega_t \) adjusts for log-nonlinearity, as
\[ \ln E_t(w_{it}) = E_t(\ln w_{it}) + \Omega_t. \]

36 In particular, we have
\[ \Psi_t \equiv \ln \left[ \Phi \left( \frac{\alpha_0 + \alpha' E(z_{it}) + \beta' \Sigma_{xz} \alpha + \sigma_{\epsilon \nu}}{\sqrt{\alpha' \Sigma_{zz} \alpha + \sigma^2_\nu}} \right) / \Phi \left( \frac{\alpha_0 + \alpha' E(z_{it})}{\sqrt{\alpha' \Sigma_{zz} \alpha + \sigma^2_\nu}} \right) \right], \]

\[ \Lambda_t \equiv \ln \left[ \frac{h_0 + \gamma \alpha_0 + \gamma \alpha' E(z_{it}) + \gamma \beta' \Sigma_{xz} \alpha + \gamma \sigma_{\epsilon \nu} + \gamma \sqrt{\alpha' \Sigma_{zz} \alpha + \sigma^2_\nu} \cdot \lambda^{\nu}_{\epsilon \nu},t}{h_0 + \gamma \alpha_0 + \gamma \alpha' E(z_{it}) + \gamma \sqrt{\alpha' \Sigma_{zz} \alpha + \sigma^2_\nu} \cdot \lambda^{\nu}_{\epsilon \nu},t} \right]. \]
Adding $\Psi_t$ adjusts for participation, as

$$\ln E_t[w_{it} | I_{it} = 1] = \ln E_t(w_{it}) + \Psi_t. \quad (95)$$

Finally, adding $\Lambda_t$ adjusts for hours-weighting, as

$$\ln \bar{w}_t = \ln E_t[w_{it} | I_{it} = 1] + \Lambda_t = E_t(\ln w_{it}) + \Omega_t + \Psi_t + \Lambda_t. \quad (96)$$

Thus, the bias expressions are complicated but the roles of $\Omega_t$, $\Psi_t$ and $\Lambda_t$ are clear. In words, the term $\Omega_t$ captures the variance of returns, observable and unobservable. The term $\Psi_t$ reflects composition changes within the selected sample of workers from which measured wages are recorded. The term $\Lambda_t$ reflects changes in the composition of hours and depends on the size of the covariance between wages and hours.

The formulation (93) of the log aggregate wage $\ln \bar{w}_t$ thus captures four important sources of variation. First, aggregate wages increase if the distribution of log-wages shifts to the right, which is the typical “well-being” interpretation of aggregate wage movements. This source is reflected by the mean $\beta_0 + \beta' \bar{x}_t$ of log-wages. Second, because individual wages are given in log form, aggregate wages will increase with increased spread of the log-wage distribution, as reflected by the heterogeneity term $\Omega_t$. Third, aggregate wages will increase if the benefit threshold increases, causing more lower-wage individuals to decide not to participate. This is reflected in the participation term $\Psi_t$. Fourth, aggregate wages will increase if the hours of higher-wage individuals increase relative to lower-wage individuals, which is captured by the hours adjustment term $\Lambda_t$. The aggregate model (93) permits estimation of these separate effects.

This framework could be relaxed in many ways. We can allow all variance terms to be time varying, as well as many of the basic behavioral parameters. If the normality assumption on the overall log-wage and participation index is not accurate for the whole population, the population can be segmented, with separate aggregate equations developed for each segment. These variations, among others, are discussed in Blundell, Reed and Stoker (2003).

4.3. Empirical analysis of British wages

The different sources of aggregate wage variation bear directly on the issue of whether aggregate wages are procyclical or not. In particular, the participation effect works counter to a normal cyclical variation of aggregate wages – decreases in participation can lead to aggregate wage increases when there is essentially no change in individual

37 Comparing $\ln \bar{w}_t$ to mean log-wage $E_t(\ln w_{it})$ is in line with the tradition of measuring “returns” from coefficients in log-wage equations estimated with individual data; c.f. Solon, Barksy and Parker (1994). Other comparisons are possible, and some may be preferable on economic grounds. For instance, if aggregate production in the economy has total human capital ($\sum_j H_j$) as an input, then the appropriate price for that input is $r_t$, so one might want to compare $\ln \bar{m}_t$ to $\ln r_t$ for a more effective interpretation. In any case, it is useful to point out that if $E_t(\ln H_j)$ is constant over time, then comparing $\ln r_t$ to $\ln \bar{m}_t$ is the same as comparing $E_t(\ln w_{it})$ to $\ln \bar{w}_t$. 


wage levels or distribution. We now turn to an analysis of British wages that shows these features.

Our microeconomic data are again taken from the UK Family Expenditure Survey (FES), for the years 1978 to 1996. The FES is a repeated continuous cross-section survey which contains consistently defined micro data on wages, hours of work, employment status and education for each year since 1978. Our sample consists of all men aged between 19 and 59 (inclusive). The participating group consists of employees; the nonparticipating group includes individuals categorized as searching for work as well as the unoccupied. The hours measure for employees in FES is defined as usual weekly hours including usual overtime hours, and weekly earnings includes overtime pay. We divide nominal weekly earnings by weekly hours to construct an hourly wage measure, which is deflated by the quarterly UK retail price index to obtain real hourly wages.

Individual attributes include education level and cohort effects. Individuals are classified into three educational groups: those who left full-time education at age 16 or lower, those who left aged 17 or 18, and those who left aged 19 or over. Dummy variables capture effects of five date-of-birth cohorts (b.1919–1934, b.1935–1944, b.1945–1954, b.1955–1964 and b.1965–1977). We include various trend variables to account for a common business-cycle effect. Finally, our measure of benefit income (income at zero hours) is constructed for each individual as described in Blundell, Reed and Stoker (2003). After making the sample selections described above, our sample contains 40,988 observations, of which 33,658 are employed, or 82.1% of the total sample.

4.3.1. Real wages and employment

Figure 8 shows log average wages in Britain from 1978 to 1996. These show a strong trend increase over the whole period. The trend appears for more disaggregate groups. Blundell, Reed and Stoker (2003) present a more detailed breakdown by cohort, region and education group, and show that the trend holds widely, including for the least-educated group.

Figure 9 shows the overall male labor employment rate for the same period. Clearly there has been a large fall in the participation rate of men. Figure 10 presents the employment rate for those with low education. For this group, there is a continued and much steeper decline in employment. This period also included two deep recessions in which there have been large fluctuations in male employment.

Considering Figures 8–10 together, one can understand the basic importance of sorting out wage growth at the individual level from changes in participation. The strong trend of aggregate wages is suggestive of great progress at increasing the well-being of

---

38 We exclude individuals classified as self-employed. This could introduce some composition bias, given that a significant number of workers moved into self-employment in the 1980s. However, given that we have no data on hours and relatively poor data on earnings for this group, there is little alternative but to exclude them. They are also typically excluded in aggregate figures.
However, great increases in unemployment are likely associated with unemployment of workers with lowest wages, or workers from the poorest groups. It is very important to understand how much of the upward trend in average wages is due to the elimination of low-wage earners from employment.

39 In fact, such a conclusion has been trumpeted by British newspapers.
There have also been well-documented changes in real benefit income over time and across different groups of individuals. While it is unlikely that variation in real value of benefit income relative to real earnings can explain all of the variation in participation rates, the changes in real benefits act as an important “instrumental variable” for separating participation decisions from determinants of wages. Again, to the extent that changes in benefit income have discouraged (or encouraged) participation, it is essential to learn the size of this impact relative to the other factors driving changes in wages.

4.3.2. Aggregation results

The Blundell, Reed and Stoker (2003) study considers a number of possible specifications for our individual-level wage equations which relate to the various specifications. In the simplest of our specifications, the full proportionality hypothesis is imposed on the (nondifferentiated) human capital model, together with trend terms to reflect the business-cycle effects on skill price. This specification was strongly rejected by the data. The preferred model had full interactions of cohort, trend, region and education. These additional variables could reflect many differences in minimum educational standards across cohorts such as the systematic raising of the minimum school-leaving age over the postwar period in the UK. The prices of different (education-level) skills are allowed to evolve in different ways, by including an interaction between high education and the trend terms. These coefficients are marginally significant and show an increasing trend among groups with higher levels of human capital. The impact of adjusting for
participation is very important. To see the impact of these results on aggregate wages, we turn to graphical analysis.

Figure 11 displays the (raw) log aggregate wage, the log aggregate wage minus the estimated aggregation bias terms, and the mean of the log wage from the selectivity-adjusted micro model. We have plotted the return from a common point at the start of the time-series rebased to zero for 1978, to highlight the changes in trend growth in wages indicated by our corrections. There is a clear downward shift in the trend, and an increased cyclical component in wage growth shown by both the corrected aggregate series and the estimated micro model.

This procedure is repeated for the lower-education group in Blundell, Reed and Stoker (2003). Several features of this analysis are worth mentioning here. For instance, even the direction of movement of the uncorrected log aggregate wage does not always mirror that of the mean micro log wage. There is a reasonably close correspondence between the two in the 1984–1988 period, but the 1990–1993 period is different. In 1990–1993, log aggregate wages are increasing, but the mean micro log wage (and the corrected aggregate wage) is decreasing – precisely the period where there is a big decline in participation. What is remarkable is that the aggregate data show reasonable growth in real wages, but such growth is virtually absent from the corrected series. We are left with a much more cyclical profile of wages.

Blundell, Reed and Stoker (2003) examined the impact of our normality assumptions by estimating with semiparametric methods. The estimated wage coefficients were hardly affected by this generalization.
If the model is exactly correct, the results from aggregating the selectivity-adjusted micro model estimates should match the corrected aggregate series. They show a close correspondence in Figure 11, and a similar close correspondence is noted by Blundell, Reed and Stoker (2003) for more disaggregated groups. In any case, we view the correspondence between the corrected log aggregate micro wage and the mean micro log wage as striking validation of the framework. This model specification that provides a good and parsimonious specification of the evolution of log real wages also seems to work well in terms of the specification of aggregation factors.

5. Conclusion

Macroeconomics is one of the most important fields of economics. It has perhaps the grandest goal of all economic study, which is to advise policymakers who are trying to improve the economic well-being of entire populations of people. In the mid-twentieth century, say 1940 to 1970, macroeconomics had an orientation toward its role much like an oracle giving advice while peering down from the top of a mountain. That is, while economists could see people making detailed decisions about buying products, investing their wealth, choosing jobs or career paths, etc., macroeconomic models were extremely simple. For instance, describing the aggregate consumption of an entire economy could be done by taking into account just a few variables: aggregate income, lagged aggregate consumption, etc. Such equations often fit aggregate data extremely well. Unfortunately, such models could not predict future aggregate variables with sufficient precision to dictate optimal policies. Even with great statistical fit, there was too much uncertainty as to what the underlying processes were that drove the aggregate data, and for policy prescriptions it is crucial to know something about those processes.

What economists could get a handle on was how rational individuals and firms would behave in various economic environments. Problems such as how to allocate one’s budget, how much to save and invest, or whether to work hard or not so hard, are sufficiently familiar that their essence could be captured with some mathematics, and

41 To get an idea of the precision of these results, Blundell, Reed and Stoker (2003) present bootstrap 95% confidence bands for the corrected log wage estimates for various groups. These plots show that the micro model prediction and the corrections to the log aggregate wage are both quite tightly estimated. In all cases, the micro model prediction and the corrections to the aggregate wage plot are significantly different from the raw aggregate wage measure and not significantly different from each other. This gives us confidence that we have identified compositional biases in the measured real wage with a reasonable degree of precision.

42 There are many stories told in the economics profession about what giants of our field thought were the greatest contributions to social science. In this spirit, we relate the following. In the mid-1980s, one of the authors asked Paul Samuelson what he felt was the greatest failure in economics. Without hesitation, his answer was “macroeconomics and econometrics”. The reason for this is that there had been an enormous anticipation in the 40s, 50s and 60s that simple empirical macroeconomic models would, in fact, be accurate enough to allow real economies to be guided and controlled, much like an automobile or a spacecraft. That this turned out to not be possible was a source of great disappointment.
economists could describe and prescribe optimal reactions. Economists could settle how someone being really smart and clear-headed would behave. Notwithstanding the anomalies pointed out recently by behavioral economists, the predictive power of economics rests on the notion that people facing a familiar situation will behave in their interests. Foolish, self-destructive or purely random behavior will not be repeated once it is consciously seen to be less good than another course. The transformation of economic analysis by mathematics occurred through the systematic understanding of rational and learning behavior by individuals and firms, and the overall implications of that for market interactions.

The merging of these two bodies of thought – macroeconomics and optimal behavior of individuals – is among the greatest developments of economics in the last half century. This advance has been recognized by Nobel prizes to Lucas, Kydland and Prescott, and one should expect more prizes to be awarded to other important developers. Previous “schools of thought” have been replaced by groups differentiated by how they settle the tradeoff between realism and strict adherence to optimal economic behavior. The specification of macroeconomic models, the judgment of whether they are sensible, and the understanding of the impacts of economic policy are now more systematic because of their embedding in the rules of optimal individual behavior.

The trouble is, this embedding cannot be right without taking account of aggregation. A one-person or five-person economy is just not realistic. One can simulate a model with a few actors and pretend that it is realistic, but there is nothing in casual observation or empirical data or economic theory that suggests that such a stance is valid. There is much to be learned from rational individual behavior, but there must be an explicit bridge to economic aggregates because real people and their situations are so very heterogeneous. Aggregation is essential, because heterogeneity is a pervasive and indisputable fact of life.

In this chapter, we have covered recent work on aggregation problems in a style that we hope is useful to empirical economists. Our orientation has been to highlight the importance of different types of individual heterogeneity: in particular, heterogeneity in tastes and reaction, heterogeneity in market participation, and heterogeneity in uninsurable risks. Our approach has been practical; we have covered recent advances in econometric modeling that address issues in aggregation, by considering explicit models at the individual level and among economic aggregates.

We have covered a wide range of ideas. First, we have detailed the main approach for incorporating distributional information into aggregate relationships, namely exact aggregation models, in the context of how that approach has been applied to the analysis of consumer demands. Second, we have shown how one can incorporate basic non-linearity, insurance and dynamic elements, in our coverage of aggregate consumption based on CRRA preferences. Third, we have shown how to account for compositional heterogeneity, in our coverage of labor participation and wages. The latter two topics required explicit assumptions on the distribution of individual heterogeneity, and we have based our solutions on normal and lognormal assumptions on individual heterogeneity. While these distributional restrictions are specific, they do permit explicit
formulations of the aggregate relationships of interest to be derived, and those formulations capture both location and spread (mean and variance) of the underlying elements of individual heterogeneity. We view our solutions in these cases as representative and clear, and good starting points for empirical modeling in the respective areas.

Whether one dates the beginning of the study of aggregation problems from the 1940s, 1930s or perhaps earlier, one can at best describe progress toward solutions as slow. Aggregation problems are among the most difficult problems faced in either the theoretical or empirical study of economics. Heterogeneity across individuals is extremely extensive and its impact is not obviously simplified or lessened by the existence of economic interaction via markets or other institutions. The conditions under which one can ignore a great deal of the evidence of individual heterogeneity are so severe as to make them patently unrealistic. There is no quick, easy or obvious fix to dealing with aggregation problems in general.

Yet we see the situation as hopeful and changing, and offer the solutions discussed in this chapter as evidence of that change. The sources of this change are two-fold, and it is worth pointing them out as well as pointing out how both are necessary.

The first source of change is the increasing availability of data on individuals observed over sequential time periods. To address questions of what kinds of individual heterogeneity are important for aggregate relationships, one must assess what kinds of heterogeneity are relevant to individual behavior for the problem at hand, and assess how much the distributions of the relevant heterogeneity vary over time. To the extent that this heterogeneity reflects differences in unexpected shocks to individual agents, the mechanisms that are available to individuals to insure against such shocks will have a strong bearing on the form of the aggregate relationship.

While we have advanced the idea of using aggregation factors (derived from time-series of individual data) to summarize the impacts of aggregation, the specific method one uses is less important than the ability to use all available types of information to study economic relationships. That is, it is important to study any relationship among economic aggregates with individual data as well as aggregate data, to get as complete a picture as possible of the underlying structure. Even though modeling assumptions will always be necessary to develop explicit formulations of aggregate relationships, testing those assumptions is extremely important, and is not possible without extensive individual data over sequential time periods. Our view is that the prospects for meaningful advance continue to brighten, as the data situation with regard to individual behavior and aggregate economic variables will continue to improve.

The second source of change in studying aggregation problems is the recent, rapid rise in computing power. Realistic accommodation of individual heterogeneity typically requires extensive behavioral models, let alone combinations of individual models with aggregate relationships. Within the last twenty five years (or the professional lives of both authors), there have been dramatic changes in the ability to implement realistic models. Before this, it was extremely difficult to implement models that are necessary for understanding impacts of individual heterogeneity in aggregation.
Aggregation problems remain among the most vexing in all of applied economics. While they have not become less difficult in the past decade, it has become possible to study aggregation problems in a meaningful way. As such, there are many reasons to be optimistic about the prospects for steady progress on aggregation problems in the future. The practice of ignoring or closeting aggregation problems as “just too hard” is no longer appropriate.

References


Further reading


Ch. 68: Models of Aggregate Economic Relationships


