Handbook of Applied Econometrics, vol 2, 1997
eds. M. Pesaran & P. Schmidt, Basil Blackwell

6

Estimating Models of Dynamic Optimization with Microeconomic Data

Robert A. Miller

1 Introduction

Dynamic considerations fraught with uncertainty are at the heart of many economic problems, and over the last 15 years a literature that explicitly accounts for them in empirical and econometric modeling has flourished. Almost all the structural modeling in this area combines tools from sequential optimization under uncertainty, with nonlinear and semiparametric estimation techniques, using data sets drawn from diverse sources. Microeconomics data, that is cross-sectional and panel data sets which rely on the cross-section for the asymptotic properties of their estimators, are playing an increasingly important role in estimating and testing dynamic structural models. This chapter reviews the class of dynamie models which have been estimated with such data and discusses the related issues of model building, identification, and statistical inference.

1.1 Discrete versus Continuous Choices

The discussion separates naturally into analyzing discrete choice models in dynamic settings, and reviewing those where choice variables come from a continuous support, because the equations characterizing the optimum differ in the two cases. When picking between a finite number of choices the value of taking each choice must be compared with the others, so optimal behavior is characterized by the same finite number of inequalities minus one. In contrast, when a choice can be varied continuously, the net effect of a marginal change on costs and benefits helps characterize the optimal decision; for example, if the problem is concave, the solution to a first-order condition fully characterizes the optimal choice. These differences in characterizing optimal behavior are reflected in modeling the econometric framework. In this respect empiricists sometimes face the question of whether the data can be most fruitfully exploited by treating the choice set as finite and identifying a person's opportunities and preferences from the payoffs from basic easily recorded choices (such as whether to enter the market or not), or focusing on choices made at the margin (such as the extent of market activity observed).

Modeling the choice set as finite implies that people with similar but not identical situations might make the same choice. Since information about their marginal differences is lost by only focusing on broad based options, the econometrician must consequently maintain stronger assumptions about the functional form of preferences, opportunities, and the nature of unobserved heterogeneity throughout the sample population to achieve econometric identification. On the other hand, focusing on the details of marginal decisions almost invites misspecification due to measurement error of various kinds. As a practical matter, it is therefore not surprising to see the same issue being addressed in one study as a discrete choice problem, yet in another as a continuous choice problem, depending on the precise nature of the data and other differentiating features of the respective approaches.

1.2 Statistical Inference and Policy Prescription

The primary virtue of estimating dynamic structural models is that their unknown parameters have a transparent interpretation within a coherent theoretical paradigm that frames the empirical investigation. Assuming the posited economic model is indeed valid, the solution to the individual's optimization problem and/or the fixed point to the equilibrium of the model, evaluated at the true unknown parameters, yields the data generating process. Therefore under the null hypothesis, the observations comprising the data set are realizations of random variables whose joint probability distribution can be formally derived from the dynamic model at the true, unknown set of parameters to be estimated. But how much must be assumed about the model in order to estimate its unknown parameters, and how will the resulting estimates guide policy analysis? Apart from deciding which choices should be modeled as continuous ones and which to treat as discrete, many dynamic models can be estimated with or without solving the optimal decision rule, thus presenting the researcher with another methodological decision.

One estimation strategy, full solution maximum likelihood (ML), is to derive the equilibrium and/or optimal decision rule for given values of the unknown parameters, in order to determine the joint probability distribution for the observed variables under the null hypothesis, and thus evaluate the likelihood function at those parameter values. This procedure is repeated for different parameter values until the vector which maximizes the likelihood is found. ML has two important virtues. Under regularity conditions usually assumed, the estimator is asymptotically efficient, attaining the Cramer–Rao lower bound in large samples. Having already computed the joint probability distribution of the observed variables at many different values, perturbing some of the estimated parameters to conduct policy analysis is straightforward.

Nevertheless the use of less efficient instrumental variables (IV) estimators that impose fewer computational burdens than ML is quite widespread. Rather than solve the model (sometimes observation by observation and parameter by parameter), these estimators exploit identifying features of the model's data generating process, such as a first-order condition for continuous choices, or an equation linking the probability of making a discrete choice to its expected payoffs, and construct sample moments from the data that converge to their population counterparts at the true parameter values. Although IV estimators are usually cheaper to compute than ML, human programming time is often required to solve the model under alternative policy regimes. But IV estimators are generally more straightforward to test, help clarify the nature of identification, and consequently encourage researchers to investigate greater numbers of more sophisticated alternatives with larger data sets. Thus ML might dominate IV where the structure is transparent (such as in engineering problems), where the data set is large enough for the finite sample properties to closely approximate their asymptotic behavior but small enough to remain computationally manageable, and where the policy proposals can clearly be cast within the structures encompassed by the parameter space. But in the face of large data tracts, and where uncertainty about the appropriate structure (such as modeling the preferences of agents over nonmarket goods) necessarily obscures the policy debate, undertaking IV estimation may be the more informative approach.

1.3 Applications

To provide a flavor of the range of applied work on dynamic discrete choice, here are some examples of empirical studies which have been undertaken, together with a brief explanation of their dynamic aspects. Within labor economics, it is convenient to arrange these studies in the order of a person's lifecycle. The economics of fertility (Wolpin, 1984; Hotz and Miller, 1988; 1993; Ahn, 1995) focuses on the income of the family, the value of the mother's time, and the demographic structure of the family (such as the number of existing children) which provide measures of the value from having an (additional) child. These studies treat offspring as a form of family capital in which choices about contraceptive measures lead to different pregnancy risks. When to quit school is the next important stage in the lifecycle that has been addressed in the literature on dynamic discrete choice (Wolpin, 1987; Tabor, 1995). The immediate benefit from taking a wage must be offset against the loss from the lower wages that less education brings. Associated with the decision to leave school is the job search process itself, reviewed by Neumann in chapter 7 of this volume; here the issue is whether a successful job applicant accepts the offer or continues to search in the hope of finding a better one. Closely related to the search process which occurs before accepting a job is the job matching and specific

human capital accumulation that occurs after a person starts working in it (Miller, 1984; Keane and Wolpin, 1994b). A person might acquire specific information and/or expertise about the current job match through working experience which only applies to that job, or opt for an alternative use of time that does not capitalize on previous work experience. Labor force (Eckstein and Wolpin, 1989b; Altug and Miller, 1997) and welfare participation (Miller and Sanders, 1996) involve similar issues. The timing of marriage and its effects on female labor supply also have dynamic implications, such as married women working fewer hours than single women (Van der Klaauw, 1993). Health care decisions and absenteeism are another dynamic problem involving trading off short leaves against a potentially more serious illness (Boswell, 1994). Finally, economists have estimated dynamic models of retirement (Berkovec and Stern, 1991; Stock and Wise, 1990; Rust and Phelan, 1995); income earned before retirement raises the standard of living in each year of retirement but reduces utility in the meantime.

Outside labor economics there has been also considerable interest in applying dynamic modeling techniques to the activities of firms. See, for example, the costs and benefits associated with renewing patents (Pakes, 1986), when to overhaul bus engines (Rust, 1987; Hotz et al., 1994), capital utilization in the cement industry (Das, 1992), job scheduling on the factory floor (Miller and Ramnath, 1995), price adjustments in the presence of menu costs (Slade, 1994), inventory policy in supermarkets (Aguirregabria, 1994), lumpy investments on farms in developing countries (Rosenzweig and Wolpin, 1993), nuclear power plant operations (Rothwell and Rust, 1995), and replacing dairy livestock (Miranda and Schnitkey, 1995). These examples give some idea of the broad range of areas that estimation techniques in dynamic structural analysis have addressed.

Perhaps the most widespread applications of continuous choices are in household consumption and individual labor supply, the former topic being extensively analyzed in chapter 4 of this volume. The discussion in this chapter emphasizes the equilibrium context in which decisions are made (Altug and Miller, 1990; Cochrane, 1991; Mace, 1991; Nelson, 1994), where a good deal of work has been done of late. For example, Townsend (1994) compares trading within and between villages; Hayashi, Altonji, and Kotlikoff (1996) investigate trading within versus between dynasties and the role of bequests in compensating for deficiencies in market structure; Miller and Sieg (1997) examine regional versus international trade; and dynamic models of equilibrium have been used to investigate taxation incidence and changing tax law (Sieg, 1995). Parallel to the rapidly growing body of work on complete markets is an earlier literature on the permanent income hypothesis (Hall and Mishkin, 1982; Zeldes, 1989; Mariger and Shaw, 1993; Luscardi, 1996). This work also brings dynamic structure to bear in estimation, although parts of it suffer from an apparent confusion between estimation with cross-section versus time series data (Chamberlain, 1984). It appears that the permanent income hypothesis does not typically impose sufficiently strong identification conditions on the budget constraint to achieve consistent estimation with a panel (Altug and Miller, 1990; Altug and Labadie, 1994; Miller and Sieg, 1997). One response to underidentification is to impose more structure on the market incompleteness (Margiotta and Miller, 1994; Rust and Phelan, 1995). Another is to relax the assumption that preferences over consumption are time additive (Hotz, Kydland, and Sedlacek, 1988; Shaw, 1989; Altug and Miller, 1997). These extensions are still in their infancy, and in time, the scope of inquiry will probably be broadened beyond labor supply and consumption.

1.4 Chapter Overview

Section 2 takes the perspective of a sample respondent.[1] It presents a canonical dynamic optimization problem from the agent's perspective, who is assumed to be solving it. The characteristic feature of such problems is that information arrives over time, and consequently decisions made later on in life depend on more information. It is convenient to partition the types of choices agents make into choices over finite sets (such as occupational choice) and those in a vector space (for example the level of consumption versus savings). Preferences are defined over the lifetime and as in almost all this literature we maintain the postulate of expected utility, which is an identifying assumption in much of the empirical work. In the case of continuous choices the solution to the agent's optimization problem often boils down to a first-order condition (called an Euler equation in dynamic problems), whereas in discrete choice problems a comparison of the utilities must be undertaken: here Bellman's (1957) principle of optimality often plays an important role.

After introducing the class of models that have been empirically investigated to date and describing the optimality conditions in the next section, sections 3 through 5 review approaches to estimating discrete choice models. One approach, described and illustrated in section 3, is to solve the agent's problem by calculating the optimal decision rule, substituting that rule back into the agent's environment, and using maximum likelihood (ML) to identify and estimate the parameters from the outcomes observed in the data. This approach is computationally intensive, and up until quite recently, apart from simple search problems with a very specialized structure (Kiefer and Neumann, 1979; 1981; Flinn and Heckman, 1982), it was believed to be the only way to estimate dynamic problems in discrete choice. Thus, most studies in this area have followed the first applications of Miller (1982; 1984) and Wolpin (1984) by solving the optimal decision rule for different sets of parameter values, and seeking the one which maximizes a likelihood function. In practice, the range of problems which can be computed using ML estimation is quite limited, and this has spurred efforts in computational algorithms to overcome the so-called curse of dimensionality. While a detailed analysis of these computational problems lies beyond the scope of this chapter, section 4 contains a brief discussion of the three important issues, namely multiple integration, algorithms for solving the

optimized lifetime utility, and approximations for the optimal decision rule. Of course, these issues have relevance within economics beyond estimation with microeconomics data, and much work on solving economic models applies to macroeconomics as well, and vice versa.[2]

As it turns out, the presumption that the optimal solution must be found to estimate dynamic models of discrete choice is unfounded. Section 5 presents a two-stage estimator which relieves the econometrician from the burden of explicitly solving the agent's discrete choice problem to obtain estimates with the usual asymptotic properties. The estimator exploits a new representation of the optimized expected lifetime utility, or the valuation function (Hotz and Miller, 1993). It is based on a relation between the conditional valuation functions which arise from making certain choices, and the probability of an outsider observing those choices being made. To explain the relation, a simple example motivates the section before the main representation result is given. Then several different estimation environments are described to show how it has been applied, while the estimation and testing issues, including a brief discussion of the small sample properties, are provided at the end of the section.

Although most discrete choice models have been estimated using ML, estimators of dynamic structural models with continuous choices have been exploiting, for many years, the first-order condition of the agent's optimization problem, thereby circumventing the problem of explicitly deriving the optimal decision rule. Much of this work stems from the following observation: if the Lagrange multiplier in a lifetime budget constraint can be recovered, estimated as a fixed effect, or differenced out of the estimation framework altogether, the Frisch demands (which are a function of prices and the marginal utility of wealth, and so are easily derived from the first-order condition when preferences are additively separable over time in the absence of impediments to markets) provide the basis for an econometric model. Early work (Heckman and MaCurdy, 1980; MaCurdy, 1981; 1983) assumed that agents had perfect foresight (that is, faced no uncertainty about the future) and could borrow or lend without constraints at a constant interest rate, and that wage growth was constant for comparable skills and demographics. Although both assumptions are soundly rejected by the data, the approach applies more generally if there are sufficient opportunities to borrow, lend, and insure (Altug and Miller, 1990). Section 6 analyzes the equilibrium generalization of the lifecycle model, including the twin assumptions of complete competitive markets and time additive preferences upon which it is based. After these assumptions are laid out and critiqued, their role in estimation and hypothesis testing is explained.

Section 7 is essentially a continuation of section 6. It analyzes three aspects of relaxing time additivity. First, it is sometimes possible to identify subeconomies, a subset of commodities and a group of traders, within which marginal rates of substitution between those commodities might be equalized for people belonging to that group. Whether this occurs because of altruism (Hayashi, Altonji, and Kotlikoff, 1996), institutional arrangements within village communities

(Townsend, 1994), or a rich set of financial and exchange rate securities (Miller and Sieg, 1997) is less important than whether resources are allocated Pareto optimally within the subeconomies themselves. Thus the first part of section 6 shows how the statistical machinery for models with the time additivity property directly applies to these subeconomies. Second, a large empirical literature has flourished purporting to test the permanent income hypothesis with microeconomics data. Here the main outstanding question is whether methods developed for time series can be imported to cross-sectional settings, a question which cannot be answered without an understanding of the equilibrium context in which agents solve their respective problems. Unfortunately, the central role of time additivity has not been properly appreciated, an issue which is therefore taken up in the second part of section 6. Then some complications associated with relaxing the other part of time additivity, additive separability of preferences, is discussed.

Finally a short conclusion in section 8 steps back from the details of this study, to provide a quick overall assessment of the advantages and disadvantages of estimating structural dynamic models with microeconomics data sets. Its purpose is not to criticize published work, although existing work is obviously the benchmark against which new contributions should be judged, but to serve as a guardrail for guiding the direction of future empirical work on dynamic structural models.

2 Dynamic Structural Models

This section develops a framework for investigating the types of dynamic optimization problems which have been analyzed empirically with micro data. First the nature of information and the resolution of uncertainty are explained. This leads into a discussion of the choices facing an agent, her preferences, as well as her production and trading opportunities. The distinction between discrete and continuous choices is emphasized, because it is reflected in the conditions that characterize optimal decision making and in the approach to estimation (discussed later in the chapter).

2.1 Information

There are several elements to the agent's dynamic problem: the information and environment the agent inhabits, how this information evolves over time, the available choices (including the times they are available), the agent's utility or objective function, and the outcomes of choices, all of which are ultimately used to rationalize behavior observed in the data. Following most of the literature in this area, this chapter assumes time progresses as a discrete variable, rather than continuously. Thus production and consumption occur on dates $t \in \{0, 1, \ldots, T\}$, where the length of the horizon T may be finite or infinite.

For all practical purposes the data econometricians work with consist of finite

dimensional vectors and a finite number of indicator variables. Consequently we will assume the model has a Markov structure, in which the agent does not need to remember all the features of the history to solve her problem, but only a summary statistic belonging to a finite vector space, here denoted $z \in Z$. Instead of keeping track of her whole history to determine her choice and its outcome, it suffices to know z_t, that is the value of the state variable z at time t, and its law of motion (meaning the probability distribution which characterizes how the state changes from period to period).

The Z space comprises anything revealed from the past that could be relevant to current and future choices or payoffs, apart from time itself, which is entered as a separate argument as t. In the special case where there is no uncertainty about the future, Z is empty. If past choices directly affect current opportunities, then these should be included within the z vector as well. (For example, previous experience in the labor market might affect current wages.) On the other hand if those choices merely increased the likelihood of an event which has occurred already (say past contraceptive practices affected the probability of pregnancy), then the event itself (whether the woman conceived or not) should enter the state space (that is, assuming children affect the household utility function), not the (contraceptive) action. Thus one component of z_t could be the number of years of schooling a woman has had, another the number of years she has held a job uninterrupted, a third might be the number of months since the most recent vacation, a fourth the size of the family, a fifth her market assets, and so on.

2.2 Choices

As mentioned above, it is convenient for model builders to partition the choices of agents into two kinds, discrete choices and continuous choices. Which choice an agent makes depends on her (current) information. Let $c_t = (c_{1t}, \ldots, c_{Mt}) \in C$ denote an M-dimensional vector of continuous choices made on date t; we require that decisions made at t only exploit information available at that time. This choice can be written as $c_t(z)$, meaning that if $z \in Z$ is the state of the world at date t, then the agent chooses $c_t = c_t(z)$, or $c_{mt} = c_{mt}(z)$ for each of the M consumption components. Specializing further, if Z is a finite set containing only L elements, then enunciating $c_t(z)$ involves stating LM values corresponding to what each consumption component would take in each state.

Information plays essentially the same role in guiding discrete choices. Suppose the agent must pick one of several actions $k \in \{0, \ldots, K\}$ in each period t. Let $d_{kt} \in \{0, 1\}$ be an indicator variable to say whether choice k is picked or not. Again we require d_{kt} to depend on z_t, the agent's information set at t. By appropriately designing the choice set we may assume without loss of generality that choices are mutually exclusive, so that the identity

$$\sum_{k=0}^{K} d_{kt} = 1 \qquad (6.1)$$

holds. This identity implies tht $d_t \equiv (d_{1t}, \ldots, d_{Kt})$ fully summarizes the agent's discrete choices, because if $d_{kt} = 0$ for all $k \in \{1, \ldots, K\}$ then $d_{0t} = 1$.

In many problems, choices are further constrained in various ways. For example, taking loans to consume large quantities of goods and services in the early periods of life may constrain consumption choices later in order to repay the debt. To incorporate these constraints, assume that for all $k \in \{0, \ldots, K\}$ there exist I real valued functions $b_{ikt}: C \times Z \rightarrow \Re$, denoted $b_{ikt}(c, z)$, which must be nonnegative. Given (c, z), each of the I index functions represents a constraint that must be met for k to be feasible. More concisely, let $b_{kt}(c, z)' = (b_{1kt}(c, z), \ldots, b_{Ikt}(c, z))$ and assume that if $d_{kt} = 1$ for (c, z) then:

$$b_{kt}(c, z) \geq 0. \tag{6.2}$$

In order to justify the Lagrangian formulation of the optimization problem used below, we will assume that if (6.2) does not define a convex set, then all the constraints are met with equality at the optimal choice. A choice $(c_t, d_t) \in C \times D$ is said to be feasible at period t for $z \in Z$ if (6.1) and (6.2) are satisfied.

2.3 Preferences

The agent's preferences are defined over the choices she makes over her lifetime, and also the state variables themselves. Because current choices affect future payoffs, current actions cannot be evaluated without stating how beliefs about the future are formed. This chapter assumes agents obey the expected utility hypothesis and models their subjective beliefs about future state variables as a probability distribution which characterizes the stochastic law of motion for z_t. When z_t is a discrete random variable (that is with finite support) this is expressed as:

$$\Pr\{z_{t+1} = z \mid z_t, c_t, d_{kt} = 1\} \equiv F_{kt}(z \mid z_t, c_t). \tag{6.3}$$

Equation (6.3) says that z_{t+1} is randomly determined by the current state variable z_t and current choices (c_t, d_t). In the case where z_t has continuous support, $F_{kt}(z \mid z_t, c_t)$ will denote the probability distribution function for z_{t+1} conditional on (z_t, c_t) and $d_{kt} = 1$, implying $z_{t+1} = z$ is replaced with $z_{t+1} \leq z$ within the probability statement on the left side of (6.3). Although this notational convention creates some ambiguity, it should not prove too confusing in what follows. In empirical applications, the rational expectations hypothesis is routinely used as an identifying assumption.[3] That is, equation (6.3) and its continuous analog are not just subjective probability distributions describing the agent's beliefs, but are also assumed to underlie the data generating process.

Given beliefs formed in this fashion, lifetime utility can be defined as the expected sum of utilities received each period. That is, for each $k \in \{0, \ldots, K\}$ and $t \in \{0, 1, \ldots, T\}$ we define a real valued mapping $u_{kt}(c_t, z_t)$ from $C \times Z$ and represent preferences by the functional

$$E\left[\sum_{t=0}^{T} \sum_{k=0}^{K} d_{kt} u_{kt}(c_t, z_t) \mid z_0\right]. \tag{6.4}$$

Expression (6.4) is interpreted as follows. Suppose that whenever the date/state coordinate pair is (t, z) the agent makes continuous choice $c_t(z)$ and discrete choice $d_t(z)$. In particular, if the agent's state at date t is z, she might choose j say, that is set $d_{jt} = 1$, and make continuous choice c_t. Then the utility she would receive that period is $u_{jt}(c_t(z), z)$. Integrating over $z \in Z$ conditional on her initial state z_0, her expected utility at time t is therefore $E[\Sigma_{k=0}^{K} d_{kt}(z) u_{kt}[c_t(z), z] \mid z_0]$. Clearly this expectation depends on all the choices and outcomes in the meantime, through the transition probability (6.3) plus the decision rules $c_s(z_s)$ and $d_s(z_s)$ for each $s < t$; hence the dynamics. Summing over periods, the lifetime expression for expected utility (6.4) is obtained.

If the only direct dependence on t comes from discounting utility received each period by a constant geometric rate over time, say $\beta \in (0, 1)$, and the horizon is infinite, meaning $T = \infty$, the optimization problem facing the agent at time s is identical to the one facing her at t providing the other state variables are the same. Therefore the optimal choices are the same, and hence independent of the date itself. Called a stationary Markov problem, it has been intensively studied in the dynamic optimization literature. In this special case $u_{kt}(c, z)$ specializes to $\beta^t u_k(c, z)$ for each $k \in \{0, \ldots, K\}$, so (6.4) simplifies to:

$$E\left[\sum_{t=0}^{\infty} \sum_{k=0}^{K} d_{kt} \beta^t u_k(c_t, z_t) \mid z_0\right]. \tag{6.5}$$

2.4 Optimization

Armed with the notation we have developed, we may state the agent's problem as follows. For each $t \in \{0, 1, \ldots, T\}$ and $z \in Z$, she picks $(c_t, d_t) \in C \times D$ to maximize (6.4) subject to (6.1) and (6.2). The solution to this problem can be interpreted either as an optimal decision rule, expressed as a function of the state z and denoted by a pair of mappings $(c_t^o(z), d_t^o(z))$, or as an optimal action conditional on the current state z_t, denoted as (c_t^o, d_t^o). Because it adequately summarizes everything the agent remembers (and her current decision cannot depend on what she has forgotten or discarded as irrelevant), this formulation of the agent's problem transmits all the dynamic factors through $F_{kt}(z^* \mid z, c)$, the transition probability which stochastically determines next period's state z^* as a function of the current period's state vector z and the choices (c, d). There are three ways z can have an effect. First, the probability distribution $F_{kt}(z^* \mid z, c)$ determining the next period's state z^* depends on the current period's state z. Second, preferences over current utility $\Sigma_{k=0}^{K} d_{kt} u_{kt}(c, z)$ may depend on z directly. Likewise, to the extent that the signs of $b_{ikt}(c, z)$ depend on z, the choice set itself is partly shaped by z.

To characterize $c_t^o(z)$ and $d_t^o(z)$, it is convenient to recast the problem recursively. To that end, the value function, $v_t(z)$ is defined for each $(t, z) \in \{0, \ldots, T\} \times Z$ by substituting the optimal decision rule back into the expected lifetime utility function:

$$v_t(z) = E\left\{\sum_{s=t}^{T}\sum_{k=0}^{K} d_{ks}^{o}(z_s)u_{ks}[c_s^{o}(z_s), z_s] \mid z_t = z\right\}. \tag{6.6}$$

Thus $v_t(z)$ is the expected remaining lifetime utility from period s onwards when the state is z and the optimal rule is implemented. Certain technical conditions must be satisfied to ensure $v_t(z)$ exists as a well defined mapping. For example, bounding the utility a person can receive in any one period suffices for finite horizon problems, and adding the assumption of discounting the future geometrically is a common way of dealing with infinite horizon optimization problems. Having calculated what to choose from period $t + 1$ onwards, making the period t choice gives the appearance of the first part of a two-period problem. Let $\lambda_{kt}(z)$ be an I-dimensional vector function conforming to the index functions $b_{kt}(c, z)$ which defines the Lagrange multiplier associated with those constraints. The agent chooses (c_t, d_t) to maximize:

$$\sum_{k=0}^{K} d_{kt}[u_{kt}(c_t, z) + \lambda_{kt}(z)b_{kt}(c_t, z) + \int v_{t+1}(z^*)\mathrm{d}F_{kt}(z^* \mid c_t, z)]. \tag{6.7}$$

This problem can be further decomposed: for each $k \in \{0, \ldots, K\}$, the agent chooses c_t to maximize:

$$u_{kt}(c_t, z) + \lambda_{kt}(z)b_{kt}(c_t, z) + \int v_{t+1}(z^*)\mathrm{d}F_{kt}(z^* \mid c_t, z). \tag{6.8}$$

Following this line of argument, let c_t^k denote the solution to maximizing (6.8), and successively substitute the solution to the K subproblems into current utility and the transition probabilities. Then define the reduced form utilities $u_{kt}(z) \equiv u_{kt}(c_t^k, z)$ and transition probabilities $F_{kt}(z^* \mid k) \equiv F_{kt}(z^* \mid c_t^k, z)$. The optimal discrete choice d_t^o maximizes:

$$u_{kt}(z) + \int v_{t+1}(z^*)\mathrm{d}F_{kt}(z^* \mid z) \tag{6.9}$$

over $k \in \{0, \ldots, K\}$, patently a finite discrete choice problem. Since equation (6.9) is often the starting point for structural estimation in discrete choice problems, it is instructive to see how a richer structural framework folds into this reduced form.

Before discussing the derivation of $v_t(z)$, several remarks are worth making about the solution to (6.8). To preface these, instead of simply writing $F_{kt}(z^* \mid c_t, z)$ for the transition probability, let us introduce a conditionally independent random variable, denoted ς, with probability density function $h_{kt}(\varsigma \mid c_t, z)$, and suppose z^* be determined by the mapping $z^* \equiv g(\varsigma, c_t, z)$, which is assumed differentiable almost everywhere. When z has continuous support, the new notation is linked to $F_{kt}(z^* \mid c_t, z)$ through the relation

$$F_{kt}(z^* \mid c_t, z) = \Pr[z \leq z^* \mid c_t, z] = \int_{g(\varsigma, c_t, z) \leq z^*} h_{kt}(\varsigma \mid c_t, z)\mathrm{d}\varsigma. \tag{6.10}$$

(In the discrete case $z \leq z^*$ is replaced with $z = z^*$ on the top line of (6.10), and $z^* \leq g(\varsigma, c_t, z)$ is replaced with $z^* = g(\varsigma, c_t, z)$ on the second line.) Assuming (6.8) is differentiable almost everywhere, c_t^k must satisfy the first-order condition (FOC)

$$0 = \frac{\partial u_{kt}(c_t, z)}{\partial c_t} + \lambda_{kt}(z)\frac{\partial b_{kt}(c_t, z)}{\partial c_t} + \int \frac{\partial v_{t+1}(z^*)}{\partial z^*}\frac{\partial g(\varsigma, c_t, z)}{\partial c_t} h_{kt}(\varsigma \mid c_t, z)\mathrm{d}\varsigma + \int v_{t+1}(z^*)\frac{\partial h_{kt}(\varsigma \mid c_t, z)}{\partial c_t}\mathrm{d}\varsigma. \tag{6.11}$$

From (6.11), there are four aspects to marginal adjustments in c_t. The effects on current utility $u_{kt}(c_t, z)$ and the shadow value of the feasibility constraints $b_{kt}(c_t, z)$ require no explanation. Continuous choices have two channels to direct their dynamic effects. The first is widely acknowledged: current choices may directly affect future utility or feasibility if lagged endogenous choices form part of the state variable. For example, cigarette consumption is widely recognized to be addictive, which is to say that the marginal utility of current consumption is an increasing function of past smoking. The second effect is often disregarded in theory and practice.[4] Adjustments in continuous choices may also affect the transition probability itself, which is the last term in expression (6.11). Continuing with the smoking example, it is also well documented that increasing cigarette consumption raises the probability of death by lung cancer and heart attack.

Equations (6.8) and (6.9) show this characterization of the optimal choices depends on the valuation function itself, yet from (6.6), the latter mapping is defined by substituting the optimal decision rule into the lifetime utility. As discussed in section 4, Bellman's (1957) principle of optimality provides a constructive way of breaking the apparent circularity of this characterization. To anticipate: supposing the horizon is finite, consider the solution to the problem in the final period T. This is a one-period static problem and therefore can be easily solved. Forming the valuation function for states z_T, the two-period problem can now be tackled. An induction completes the construction for T. When future utility is discounted, the infinite horizon case can be approximated arbitrarily closely by a finite horizon problem of sufficient length, and this provides the key to the proof, which is based on an application of the contraction mapping theorem.

3 ML Estimation of Discrete Choice Models

Given laws of motion for production and preferences, the optimal choices made in equilibrium generate the stochastic processes that describe the variables econometricians observe in their cross-sectional or panel data set. The probability distributions associated with these stochastic processes form the basis for undertaking statistical inference. At the heart of structural estimation is the notion that the criterion function for an estimator of the unknown parameter vector should be explicitly derived from the stochastic process that helps characterize optimal or equilibrium behavior in the theoretical model itself. The next two sections of this chapter explore this idea for dynamic discrete choice models. From a conceptual viewpoint, maximum likelihood (ML) estimation is the most straightforward to discuss, and is therefore presented first.

3.1 Variables and Parameters

To date, all empirical work on dynamic discrete choice has concentrated on models where everything, except the values of a finite number of parameters, is known about the model.[5] This deficiency bespeaks another, the problem of identification, both within the parametric structure imposed by respective model builders (and the associated computational difficulties encountered in estimation), and more broadly with respect to alternative economic explanations. Since one major goal of structural estimation is to recover policy invariant structure in order to provide a testbed for numerically analyzing alternative policies, these two related limitations are particularly galling. Unfortunately, policy recommendations from such models often appear sensitive to exactly what is assumed about the parametric structure being imposed. Nevertheless the alternative research strategy of not interpreting estimated coefficients within any properly articulated economic paradigm seems even less attractive, and that indictment has continued to propel the advocates of structural econometrics. Having plainly acknowledged these misgivings, this chapter resumes the business of analyzing what has been achieved to date. We start with the standard assumption that the econometrician's goal is to estimate the unknown parameter vector $\theta_0 \in \Theta$ in a model that is fully specified up to Θ, a compact convex subset of a Euclidean vector space.

If all the variables affecting decisions and outcomes were retrospectively observed, armed with the true values of the parameters, the economist could deduce at each point in time exactly how optimizing agents should behave, and the precise consequences of those decisions. Consequently any deviation between the model and the data it generated could be attributed solely to misspecification. Similarly, if only a finite number of parameters are unknown, then typically an equally small number of observations are required to infer their true values, and the remaining observations can be used to check for specification errors, a single deviation being sufficient to reject the framework. Because economists never attain such dramatic precision in forecasting, unobserved variables (including measurement error) are introduced to reconcile their models with the data.

3.2 Estimation

The statistical properties of the model critically depend on just where the unobserved variables enter. Following section 2, suppose the model has a state space representation and, for a randomly selected sample of people, choices are observed for (at least) one period, along with all the state variables both immediately before and after the choice is made. Changing the subscripting convention from the previous section to now reflect the orientation towards estimation with a cross-section or panel, let (d_n, z_n, z_n^*) now denote for each person $n \in \{1, \ldots, N\}$ in the sample; her decision d_n at the time she was interviewed, her state variable when making the decision z_n, and the state variable for next period z_n^*. Then the only reason for not producing an exact fit could be that the law of motion for the transition equation is stochastic. So if θ_0 is the unknown parameter to be estimated, then in the special case where z_n is one of L elements, $F_k(z_n^* \mid z_n, \theta_0)$ denotes the probability that z_n^* occurs next period given z_n and, say, that n chooses action k. If $F_k(z_n^* \mid z_n, \theta_0)$ is degenerate, meaning z_n^* is fully determined by z_n and k, then θ_0 is fully revealed by a small number of observations. But assuming $F_k(z_n^* \mid z_n, \theta)$ is a proper probability distribution for all $\theta \in \Theta$, and the data are a random sample, the ML estimator for the transition probability is found by maximizing the log-likelihood

$$\sum_{n=1}^{N} \ell_n(\theta) \tag{6.12}$$

with respect to $\theta \in \Theta$, where $\ell_n(\theta)$ is defined in the case where z_n has discrete support as:

$$\ell_n(\theta) = \sum_{k=0}^{K} d_{kn} \ln\,[F_k(z_n^* \mid z_n, \theta)]. \tag{6.13}$$

The large sample properties of the ML estimator are derived in the usual way: it is consistent, it converges in the square root of the sample size, and its asymptotic covariance matrix attains the Cramer–Rao lower bound.[6] If, however, all the state variables and choices are observed, then the decision rule is fully determined from knowledge about the true parameter values. Thus by a similar argument to above, one could deduce the agent's decisions (exactly) from a small number of observations, subject to the all-important caveat that the model is correctly specified.

To avoid such stochastic singularities, researchers introduce unobserved state variables into their formulations. For exposition purposes, let us now assume that the transition law $F_k(\cdot \mid z_n)$ is known, perhaps because it is deterministic or has already been estimated, and that an unobserved state variable enters the model. This pragmatic treatment of unobserved variables, namely introducing unobserved variables to apply the formalisms of statistical inference, contrasts with the more idealistic notion that the chief virtue of a structural model is that it represents the true model.[7] Here the objective appears more modest: to present micro data in a parsimonious way, so that the parameters have a clean interpretation within economics, and that no hypothesis tests derived from the model are rejected by the data.[8]

In this spirit of empiricism, partition the state variable for observation n into its observed and unobserved components. Thus $z_n \equiv (x_n, \varepsilon_n)$, where x_n is observed and ε_n is unobserved. The decision rule characterized in section 2 can then be expressed as $d(x_n, \varepsilon_n, \theta_0)$ for some unknown $\theta_0 \in \Theta$, and the probability that the econometrician observes $n \in \{1, \ldots, N\}$ taking action $k \in \{0, \ldots, K\}$ is:

$$\Pr\{d_k(x_n, \varepsilon_n, \theta_0) = 1 \mid x_n\} = \int d_k(x_n, \varepsilon_n, \theta_0) dG(\varepsilon_n \mid x_n, \theta_0), \tag{6.14}$$

where $G(\varepsilon_n \mid x_n, \theta_0)$ is the probability distribution for ε_n conditional on x_n. Given computationally tractable methods of computing $d(x_n, \varepsilon_n, \theta)$ and $G(\varepsilon_n \mid x_n, \theta)$ for each sample point $n \in \{1, \ldots, N\}$ and $\theta \in \Theta$, from (6.13) the ML estimator for θ_0 maximizes (6.12) with $\ell_n(\theta)$ now defined as:

$$\ell_n(\theta) = \sum_{k=0}^{K} d_{kn} \ln \{\int d_k(x_n, \varepsilon_n, \theta) dG(\varepsilon_n \mid x_n, \theta)\}. \tag{6.15}$$

The estimator has the same desirable asymptotic properties mentioned above.

3.3 An Example

To illustrate the principle of ML estimation in dynamic models of discrete choice, consider the following job matching model (Jovanovic, 1979; Miller, 1984; Flinn, 1986; Keane and Wolpin, 1994a; Flyer, 1995). In period t a worker receives compensation from her current job, denoted λ_t, which is the sum of two terms, a fixed component ξ, called the value of a job match which only changes with job turnover, and a transitory component η_t which is common across matches:

$$\lambda_t = \xi + \eta_t. \tag{6.16}$$

Each period she chooses between taking a new job and remaining in her current position. If the agent knows the value of job matches without gaining any specific experience, then she accepts a new job with fixed component denoted ξ^* if and only if ξ^* exceeds the value of the current job match ξ. In that case the dynamic features of the problem are trivial.

But suppose she sees neither the transitory component η_t nor the value of a job match ξ, and only λ_t is revealed at the end of period t to the worker. It is assumed that ξ^* is drawn from a normal distribution with mean γ and variance δ^2, and also that η_t is an identically and independently distributed normal random variable over t with mean 0 and variance σ^2. In this case the agent has an extra reason for sampling a new job (apart from current compensation): to acquire information about the value of her job match. By changing the features of the information structure, a trivial problem becomes quite challenging.

The state variables for the (harder) problem are the posterior beliefs of the job the person worked in last period. Let γ_t denote the posterior mean of the agent's posterior distribution for the current job match at time t, and δ_t^2 its variance. Appealing to Bayes' rule, it follows that the agent updates her prior beliefs each period as she receives the return λ_t from her current match. More specifically, let $d_t = 0$ if the worker takes a new job in period t, with $d_t = 1$ otherwise. Defining the information factor $\alpha_t \equiv \sigma^2\delta_t^{-2}$ as a noise to information ratio, γ_t and α_t follow the Markov process

$$\gamma_{t+1} = d_t(\alpha_t\gamma_t + \lambda_t)/(\alpha_t + 1) + (1 - d_t)(\alpha_0\gamma_0 + \lambda_t^*)/(\alpha_0 + 1) \tag{6.17}$$

$$\alpha_{t+1} = d_t(\alpha_t + 1) + (1 - d_t)(\alpha_0 + 1), \tag{6.18}$$

where λ_t^* is the first period's return in a new job.[9] This is a stationary Markov problem. In terms of section 2 notation, the vector of state variables reduces to $z_t \equiv (\gamma_t, \alpha_t)$. Following the notational convention adopted in (6.5), current utility expected at the beginning of the period as a function of choice can be expressed as $u_1(z) = \gamma_t$ and $u_0(z) = \gamma_0$. Continuing in this manner, let $v_1(\gamma_t, \alpha_t)$ be the conditional valuation function for remaining on the current job with characteristics (γ_t, δ_t) for one more period (and behaving optimally thereafter):[10]

$$v_1(\gamma_t, \alpha_t) = \gamma_t + E\left[\sum_{s=t+1}^{\infty} \beta^s [d_s^o\gamma_s + (1 - d_s^o)\gamma_0] \mid \gamma_t, \alpha_t, d_t = 1\right]. \tag{6.19}$$

When the agent takes a new job, she confronts her initial conditions again. As is true for all stationary renewal problems (of which this is an example), the conditional valuation function for $d_t = 0$ simplifies to a constant, here $v_0 \equiv v_1(\gamma_0, \alpha_0)$; and d_t^o, the optimal decision rule, is defined by the recursion

$$d_t^o = \begin{cases} 1 & \text{if} \quad v_1(\gamma_t, \alpha_t) > v_0 \\ 0 & \text{if} \quad v_1(\gamma_t, \alpha_t) \le v_0. \end{cases} \tag{6.20}$$

Therefore the valuation function is:

$$v(\gamma_t, \alpha_t) = \max\{v_1(\gamma_t, \alpha_t), v_0\}. \tag{6.21}$$

Turning now to the estimation aspects, suppose there are data on job turnover, but not returns (which might have a nonpecuniary component). More specifically the data consist of N observations on job matches from an economy described by $(\beta, \gamma_0, \delta_0, \sigma)$, in which $\tau(n)$, the length of the match, is recorded along with an indicator variable, d_n, saying whether the match has ended, in which case $d_n = 0$, or not ($d_n = 1$). Because λ_t is unobserved, only (α_0, β), rather than the fourtuple $(\beta, \gamma_0, \delta_0, \sigma)$, is identified (Miller, 1984). Let $q_\tau(\alpha_0, \beta)$ denote the probability that a job match ends after τ periods. Then the probability that a spell remains incomplete after ρ periods is $[1 - \Sigma_{s=1}^{\rho} q_s(\alpha_0, \beta)]$. Define $\theta_0 = (\alpha_0, \beta)$. Then the ML estimator is found by maximizing (6.12) with respect to θ, where:

$$\ell_n(\theta) = d_n \ln\left[1 - \sum_{s=1}^{\tau(n)} q_s(\alpha, \beta)\right] + (1 - d_n) \ln \{q_{\tau(n)}(\alpha, \beta)\}. \tag{6.22}$$

In principle it is straightforward to extend this ML estimator to heterogeneous populations. If the heterogeneity is attributable to a vector of observed variables x_n, then α and β could be parameterized by known functions $\alpha(x_n, \theta)$ and $\beta(x_n, \theta)$, and substituted into (6.22) for the maximization of (6.12) over $\theta \in \Theta$. If differences throughout the population are unobserved, then these must be integrated out. Let $H(\varepsilon, \theta_0)$ be the probability distribution for ε_n, an unobserved random variable. Then the probability of observing a match severed at τ periods is:

$$q_\tau(\theta) = \int q_\tau(\alpha(\varepsilon, \theta), \beta(\varepsilon, \theta), \theta)\mathrm{d}H(\varepsilon, \theta). \qquad (6.23)$$

Substituting $q_\tau(\theta)$ for $q_\tau(\alpha, \beta)$ in (6.22) and maximizing over θ, the ML estimator is obtained for this case too.

This example demonstrates that many phenomena econometricians encounter in other estimation contexts also arise in the estimation of structural dynamic models. Right censoring (due to the fact that current spells are incomplete), choice based sample selection (which occurs only because job matches with better histories survive), controlling for observables (to deal with the effects of education on the information factor), and integrating out random effects (because not all the differences among agents are observed) are familiar terms to researchers working with cross-sectional data sets.

4 Computation

There is one further feature that applications of ML to structural dynamic models share with many other nonlinear ML estimation environments, namely the high cost of implementation. The computational burden of obtaining estimates for the job matching model comes from two sources. First is the problem of computing the value function many times (as the estimation program tries out different values of the unknown parameters within the likelihood function). The second source of the computational burden comes from the numerous integrations required to implement the ML estimator when unobserved state variables exhibit time dependence. For example one of the state variables in the job matching model not usually recorded in the data, a person's belief about the current job match, evolves as a martingale. This section briefly reviews algorithms used to compute the estimator, and the attempts to surmount its perplexing computational aspects.

For two reasons this chapter does not explicitly deal with ML estimators for fully solved continuous choice dynamic models. First, notwithstanding promising recent work on estimating continuous choice models using full solution methods (Miranda and Rui, 1996), the vast majority of dynamic continuous choice models are estimated from restrictions on the data implied by the first-order conditions, rather than the value function itself. Second, much of the analysis undertaken here for discrete models also directly applies to continuous models; indeed one method for handling a continuous choice model is to first

4.1 Integration and Simulation

Multiple integration within the likelihood function, to account for serial correlation in unobserved parts of the state variable, is not unique to choice theoretic frameworks, but shared with many limited dependent variable estimation settings where dynamics are important (Heckman, 1981a; 1981b). To illustrate the problem in a dynamic programming context, let us reconsider the job matching model discussed in the previous section. Let $G(\gamma \mid \alpha)$ denote the probability distribution for γ across the population of workers whose information factor about their current job match is α. As tenure increases, the population of job matches is winnowed by Bayesian updating and selection. Since everybody of a given type starts with common prior beliefs about their suitability to a new job match, $G(\gamma \mid \alpha_0)$ is a degenerate distribution with a single mass point at γ_0. With one draw from a period's experience on the job, the population distribution of mean beliefs fans out. Those people who receive the worst draws quit at once, thus truncating the dispersed distribution from below; the remainder survive with their updated beliefs to take at least one more draw and an extra period's job experience. Thus $G(\gamma \mid \alpha_0 + 1)$ is truncated below at the reservation mean belief for quitting after one period, and is continuously distributed over the real line above the truncation point. This process is repeated indefinitely.

The value function $v(\gamma, \alpha)$ is increasing in γ. Hence, from equation (6.20), a worker with information factor α quits at the beginning of the period if $\gamma < \tilde{\gamma}(\alpha)$, where $\tilde{\gamma}(\alpha)$ is the unique solution to $v_0 = v_1[\tilde{\gamma}(\alpha), \alpha]$. Thus the probability of quitting the match with tenure τ, that is conditional on accumulating τ period's experience, is $G[\tilde{\gamma}(\alpha_0 + \tau) \mid \alpha_0 + \tau]$, which is called the hazard rate. The unconditional probability of a job match lasting exactly τ periods, denoted q_τ, can be calculated from the hazard rates. It is the product of the probability of surviving τ periods, $\prod_{s=1}^{\tau-1}\{1 - G[\tilde{\gamma}(\alpha_0 + s) \mid \alpha_0 + s]\}$, and the hazard from quitting at that point, $G[\tilde{\gamma}(\alpha_0 + \tau) \mid \alpha_0 + \tau]$:

$$q_\tau = G[\tilde{\gamma}(\alpha_0 + \tau) \mid \alpha_0 + \tau] \prod_{s=1}^{\tau-1} \{1 - G[\tilde{\gamma}(\alpha_0 + s) \mid \alpha_0 + s]\}. \qquad (6.24)$$

Consider the steps required to solve $G(\gamma \mid \alpha)$ numerically. As mentioned in the previous section, only the information and discount factors (α_0, β) are identified. So without loss of generality γ_0 may be set to 0 and σ to 1. Since $G(\gamma \mid \alpha_0 + 1)$ is just the probability that $\gamma_1 \leq \gamma$, it immediately follows from the Bayes updating rule (6.17) that $\gamma_1 = \lambda_0/(\alpha_0 + 1)$. Using (6.16) to substitute out λ_0:

$$G(\gamma \mid \alpha_0 + 1) = \Pr\{\xi + \eta_0)/\alpha_0 + 1) < \gamma\}. \qquad (6.25)$$

Because the match parameter ξ is initially distributed as a normal random variable with mean 0 and variance α_0^{-1}, while η_0 is an independent standard normal random variable, workers with tenure of one period view $(\xi + \eta_0)/(\alpha_0 + 1)$ as a normal random variable with mean 0 and variance $[\alpha_0(\alpha_0 + 1)]^{-1}$. Therefore:

$$G(\gamma \mid \alpha_0 + 1) = \Theta[\gamma\sqrt{(\alpha_0(\alpha_0 + 1))}] \qquad (6.26)$$

and $\Phi[\bar{\gamma}(\alpha_0 + 1)\sqrt{(\alpha_0(\alpha_0 + 1))}]$ is the hazard rate for workers quitting after one period. Clearly $G(\gamma \mid \alpha_0 + 1)$ can be found by calling standard routines for the normal cumulative distribution function.

Those with tenure $\tau > 1$ update γ_τ to $\gamma_{\tau+1}$ using (6.17), but now the integration is not only over the probability distribution for η_τ, but also over γ_τ. Denoting $\alpha_0 + \tau$ by α and using the same reasoning as above:

$$G(\gamma \mid \alpha + 1) = \Pr\{\gamma_{\tau+1} < \gamma \mid d_{\tau+1} = 1\}$$
$$= \Pr\{(\alpha\gamma_\tau + \xi + \eta_\tau)/(\alpha + 1) < \gamma \mid d_{\tau+1} = 1\}, \tag{6.27}$$

where the conditioning statement on the right side of (6.27) is explicit about considering only those matches that last more than τ periods. Integrating over job matches with γ_τ who meet this survival criterion:

$$G(\gamma \mid \alpha + 1) = E\{\Phi[(\gamma - \gamma_\tau)\sqrt{(\alpha(\alpha + 1))}]\}$$
$$= \{1 - G[\bar{\gamma}(\alpha) \mid \alpha]\}^{-1}\left\{\int_{\bar{\gamma}(\alpha)}^{\infty} \Phi[(\gamma - \gamma_\tau)\sqrt{(\alpha(\alpha + 1))}]dG(\gamma_\tau \mid \alpha)\right\}. \tag{6.28}$$

Substituting the normal probability density associated with (6.26) into (6.28) to determine $G(\bar{\gamma}(\alpha_0 + 2) \mid \alpha_0 + 2)$, it is clear that the integration must be undertaken numerically. Similarly the probability density for $G(\bar{\gamma} \mid \alpha_0 + 2)$ can only be computed using numerical methods. Continuing in this manner, the algorithm recursively determines $G[\bar{\gamma}(\alpha_0 + \tau) \mid \alpha_0 + \tau]$ for any finite length τ sequence as required by (6.24), the approximation error accumulating with the length of the sequence.[11] Naturally the whole procedure must be executed on each candidate $\theta \in \Theta$ when maximizing the log-likelihood (6.22).

Rather than numerically integrate over the distribution of unobserved variables, simulating a supplementary pseudo data set, proportional in size to the actual one, provides the basis for an alternative estimation strategy. Sample moments generated by the simulations (interpreted as artificially generated data) are equated to corresponding sample moments from the real data by adjusting the underlying parameters to be estimated. In the job matching example, the pseudo data could be generated as follows. Recalling N is the number of observations, draw, say, six outcomes from the standard normal distribution (if the aim is to estimate the model from the first five decision periods of job experience) for each observation, that is $6N$ random numbers in total, denoted $(\varepsilon_{n1}, \ldots, \varepsilon_{n6})$ for all $n \in \{1, \ldots, N\}$. Given a value for α_0, and assuming as before $\bar{\gamma}(\alpha)$ is known, let $\alpha^{-1/2}\varepsilon_{n1}$ denote the match parameter ξ_n associated with this simulated match, and define $\lambda_{ns} \equiv \alpha_0^{-1/2}\varepsilon_{n1} + \varepsilon_{n,s+1}$ as the return in the sth period of tenure. Using (6.17) update beliefs about ξ_n; thus the posterior mean for ξ_n with one period's experience is $(\alpha_0\varepsilon_{n1} + \varepsilon_{n2})/(\alpha_0 + 1)$. If this number is less than $\bar{\gamma}(\alpha_0 + 1)$, then the simulated match ends, the event being denoted $d_{n1}(\alpha_0, \beta) = 1$. Otherwise the match survives at least one more period, $d_{n1}(\alpha_0, \beta = 0$, and another return is drawn. Continuing in this manner, the artificial variables $d_{ns}(\alpha_0, \beta) \in \{0, 1\}$ are constructed for $s \in \{1, \ldots, 5\}$; the vector function

$$d_n(\alpha_0, \beta) = (d_{n1}(\alpha_0, \beta), \ldots, d_{n5}(\alpha_0, \beta))' \tag{6.29}$$

is formed, averaged over $n \in \{1, \ldots, N\}$, and matched against the corresponding sample moment from the data. More specifically, set $d_{n\tau} = 1$ if the nth person in the sample quits with τ periods tenure, and set $d_{n\tau} = 0$ otherwise. Forming the vector $d_n = (d_{n1}, \ldots, d_{n5})'$, a consistent and asymptotically normal estimator for the unknown true value of $\theta \equiv (\alpha_0, \beta)$, denoted $\theta_0 \in \Re^+ \times [0, 1]$, is obtained by minimizing:

$$\left[N^{-1}\sum_{n=1}^{N}(d_n - d_n(\theta))\right]' A_N \left[N^{-1}\sum_{n=1}^{N}(d_n - d_n(\theta))\right] \tag{6.30}$$

with respect to $\theta \in \Re^+ \times [0, 1]$, where A_N converges to a symmetric matrix of full rank A.

The asymptotic covariance matrix for the estimator takes the standard form $(D'AD)^{-1}D'A\Sigma A'D(D'AD)^{-1}$, where:

$$\Sigma = E[(d_n - d_n(\theta_0))(d_n - d_n(\theta_0))']$$
$$D = E\left[\frac{\partial q_n(\theta_0)}{\partial \theta}\right] \tag{6.31}$$

and $q_n(\theta_0)$ is the probability of the sampled person quitting when they are observed to leave in the data (McFadden, 1989; Pakes and Pollard, 1989). However the proofs of consistency and asymptotic normality are complicated by the fact that the criterion function (6.30) is not continuous in the parameters: as $\theta \equiv (\alpha_0, \beta)$ varies a little, each indicator function $d_{ns}(\alpha_0, \beta)$ remains the same or switches to its alternative value. Intuitively, the discontinuity induced by the contribution of each observation diminishes in importance as the sample size grows, in an analogous way to the graph of the cumulative relative frequency of any real value random variable with a well defined probability density, a step function, converging to its smooth underlying probability distribution function. Rather than draw only one set of random numbers $(\varepsilon_{n1}, \ldots, \varepsilon_{n6})$ for each observation $n \in \{1, \ldots, N\}$ in the data, a total of R sets could be drawn, labeled $(\varepsilon_{1n}^{(r)}, \ldots, \varepsilon_{6n}^{(r)})$, from which a sequence of R durations for each artificial observation $n \in \{1, \ldots, N\}$ could be generated using the same procedure as above. In this case $d_n(\alpha_0, \beta)$ would be defined as:

$$d_n(\alpha_0, \beta)' = \left(R^{-1}\sum_{r=1}^{R} d_{n1}^{(r)}(\alpha_0, \beta), \ldots, R^{-1}\sum_{r=1}^{R} d_{n5}^{(r)}(\alpha_0, \beta)\right). \tag{6.32}$$

The advantage of simulating more than a single hypothetical job match duration per observation is asymptotic efficiency; Σ, and hence the covariance matrix, adjusts by a factor of $(1 + R^{-1})$. By the law of large numbers, $d_n(\alpha_0, \beta)$

converges to $q_n(\alpha_0, \beta)$ for each (α_0, β) as R increases indefinitely, thus illustrating how Monte Carlo simulation could be used to undertake the required numerical integrations.

Although the asymptotic distribution is hardly affected by this clever device for large but finite R, two cautionary remarks are offered; both question how well the asymptotic correction in the standard errors reflects the errors generated by the simulation procedure. The first is that the sample size required for the finite distribution to approximate its asymptotic counterpart may be quite large. Second, the larger RN (from increasing the number of simulations per observation or increasing the data set), the more difficult it seems to be to generate longer sequences of random numbers to serve as hypothetical. After all, these numbers are not randomly generated, but come from computer generated outcomes of a deterministic process.[12]

4.2 Solving the Value Function

Integrating over future paths is only one of the computational problems confronted by researchers estimating dynamic models with ML. A second issue is how to compute the value function itself. When the likelihood is differentiable, the effect of small changes in the parameters can be well approximated by derivative based methods, which reduces the computational burden as in any nonlinear optimization problem. Nevertheless, this still leaves many evaluations to be undertaken by directly calculating the value function at many different parameter values within the likelihood at each set of state variables observed in the sample, a very costly undertaking for models with moderately large state spaces and several parameters to estimate.

Perhaps the most natural way to look at finite horizon dynamic programming is from Bellman's (1957) perspective of backwards induction mentioned in section 2, and many microeconometric applications use this method (such as Wolpin, 1984). Continuing the job matching example, and as before normalizing with $\gamma_0 = 0$ and $\sigma = 1$, suppose the problem has a finite horizon T, and consider the choice facing a worker entering the last period with state variables (γ_T, α_T). Her value function in period T is simply $v_T(\gamma_T, \alpha_T) = \max\{\gamma_T, 0\}$. Taking the expectation of $v_T(\gamma_T, \alpha_T)$ one period before when her state variables are (γ, α) yields:

$$\begin{aligned} g_{T-1}(\gamma, \alpha) &\equiv E[v_T(\gamma_T, \alpha_T) \mid \gamma, \alpha] \\ &= E[\max\{(\alpha\gamma + \lambda_{T-1})/(\alpha + 1), 0\}] \\ &= \gamma - \gamma\Phi[-\gamma\surd(\alpha(1 + \alpha))] + [\alpha(1 + \alpha)]^{-1/2} \int_{-\gamma\surd(\alpha(1+\alpha))}^{\infty} \varepsilon d\Phi(\varepsilon), \end{aligned} \tag{6.33}$$

where $\Phi(\varepsilon)$ denotes the cumulative distribution function for a standard normal random variable. The third line in (6.33) is derived as follows. Using the arguments preceding equation (6.30), the probability that the person will quit at the beginning of the next period is $\Phi[-\gamma\surd(\alpha(1 + \alpha))]$, in which case her expected return next period is 0. Should she remain with her current match, which occurs with probability $1 - \Phi[-\gamma\surd(\alpha(1 + \alpha))]$, her expected return as of the beginning of the next period is γ plus the expected change in this random variable:

$$\{[1 - \Phi[-\gamma\surd(\alpha(1 + \alpha))]\surd(\alpha(1 + \alpha))\}^{-1} \int_{-\gamma\surd(\alpha(1+\alpha))}^{\infty} \varepsilon d\Phi(\varepsilon). \tag{6.34}$$

Collecting terms, the bottom line in (6.33) is obtained.

Having calculated $g_{T-1}(\gamma, \alpha)$, at the beginning of period $T - 1$ the person chooses the maximum of starting a new job, and continuing with the current one:

$$v_{T-1}(\gamma, \alpha) = \max\{\gamma + \beta g_{T-1}(\gamma, \alpha), \beta g_{T-1}(0, \alpha_0)\}. \tag{6.35}$$

Now taking the expectation of the expected value of $v_{T-1}(\gamma, \alpha)$ with respect to (γ, α) conditional on the state variables $(\gamma_{T-2}, \alpha_{T-2})$ at date $T - 2$, the expected valuation function $g_{T-2}(\gamma_{T-2}, \alpha_{T-2})$ is obtained. By successively solving for the functions $g_{T-1}(\gamma, \alpha)$ through $g_1(\gamma, \alpha)$, the value function $v_t(\gamma, \alpha)$ is derived numerically for all $t \in [0, \ldots, T]$. It is easy to see from this description that numerical error accumulates with each iteration.

The contraction mapping theorem extends the idea of iterating on the value function to infinite horizon settings, and thus provides a numerical algorithm for solving them (as applied by Miller, 1984 and others). The contraction mapping theorem is based on the idea that if the next period's expected utility is bounded (with arbitrarily high probability) and the discount rate is strictly less than one, then the amount of utility lost from ignoring distant events altogether is less than the sum of the tail of a geometric series.[13] Turning to the job matching example, let $h_0(\gamma, \alpha)$ be any real value bounded continuous function defined on the coordinates (γ, α) and define the real valued mapping $C[h(\gamma, \alpha)]$ as:

$$\begin{aligned} C[h(\gamma, \alpha)] \equiv E[&\max\{\gamma + \varepsilon/\surd(\alpha(1 + \alpha)) \\ &+ \beta h(\gamma + \varepsilon/\surd(\alpha(1 + \alpha)), \alpha + 1), \beta h(0, \alpha_0)\}], \end{aligned} \tag{6.36}$$

where the integration is over ε, a standard normal random variable. From the reasoning used to justify equation (6.35), $C(\cdot)$ recursively links successive expected value functions in the finite horizon problem with $g_k(\gamma, \alpha) = C[g_{k-1}(\gamma, \alpha)]$. Mappings such as $C(\cdot)$ meeting the conditons of the theorem are called contractions, and satisfy the fixed point property that a unique $g(\gamma, \alpha)$ solves $g(\gamma, \alpha) = C[g(\gamma, \alpha)]$. In this case $g(\gamma, \alpha)$ is interpreted as the expected value function for the infinite horizon problem:

$$g(\gamma, \alpha) = E[v(\gamma + \varepsilon/\surd(\alpha(1 + \alpha)), \alpha + 1) \mid \gamma, \alpha]. \tag{6.37}$$

A corollary to the contraction mapping theorem gives an upper bound to the distance between the value function and an iterate in the sequence of approximating functions induced by $C(\cdot)$. In particular for any initial $h_0(\gamma, \alpha)$, the normed distance between $C^t[h_0(\gamma, \alpha)]$ and $g(\gamma, \alpha)$ is less than the normed distance between the successive iterates $C^t[h_0(\gamma, \alpha)]$ and $C^{t+1}[h_0(\gamma, \alpha)]$ scaled up by the factor $(1 - \beta)^{-1}$. Notice that in contrast to the recursion used to numerically derive $G(\gamma \mid \alpha)$, and in contrast to finite horizon value functions discussed above, numerical errors do not accumulate when the contraction mapping is used in an infinite horizon problem because the approximation error only depends on the final two iterates. Although convergence is global, an intelligent choice of the initial function $h_0(\gamma, \alpha)$ reduces the number of iterations required to achieve convergence. One such choice might be:

$$h_0(\gamma, \alpha) = (1 - \beta)g_{T-1}(\gamma, \alpha), \tag{6.38}$$

which is the discounted lifetime utility of updating the current match for one more period and then choosing the maximum of starting a new match every period versus never leaving the current match.

An alternative to value function iteration is policy function iteration (which Rust, 1987 used). Like value function iteration but unlike the finite horizon case, the numerical errors do not accumulate. Under the policy iteration alogrithm an initial rule is picked, say $d_1(\gamma, \alpha)$. The infinite lifetime expected value of applying this rule, denoted $w_1(\gamma, \alpha)$, is then calculated as:

$$w_1(\gamma, \alpha) = E\left[\sum_{t=0}^{\infty} \beta^t d_1(\gamma_t, \alpha_t)\lambda_t \mid \gamma, \alpha\right], \tag{6.39}$$

and $d_2(\gamma, \alpha)$ as the implicit solution to $C[w_1(\gamma, \alpha)]$, maximizing:

$$E[\gamma + \varepsilon/\sqrt{(\alpha(1 + \alpha))} + \beta w_1(\gamma + \varepsilon/\sqrt{(\alpha(1 + \alpha))}, \alpha + 1), \beta w_1(0, \alpha_0) \mid \gamma, \alpha]. \tag{6.40}$$

The infinite lifetime expected value $w_2(\gamma, \alpha)$ is calculated under $d_2(\gamma, \alpha)$, and so forth. Since $w_k(\gamma, \alpha)$ is, by inspection, increasing in k, bounded above by the unique fixed point $g(\gamma, \alpha) = C[g(\gamma, \alpha)]$, it follows that the $w_k(\gamma, \alpha)$ sequenee converges to $g(\gamma, \alpha)$.

A policy function iteration entails substantially more work than the maximization step (6.35) which defines a value function iteration. The additional step (6.39) is computationally demanding unless the state space for the problem is finite and small. The benefit from taking the solution step within each value function iteration is that convergence is achieved in fewer overall iterations. Intuitively each value iteration tacks one period to the horizon on the front of the agent's lifetime problem, and does not update the rules that are used in future periods; consequently decisions in the finite horizon program remain time dependent. In contrast policy iteration fixes the length of the horizon at infinity, and exploits the feature that the optimal decision rule is stationary, by restricting the algorithm to search over a class of stationary rules.

4.3 Approximating Optimality

Because of the computational expense involved, several studies in the microeconometric literature on discrete choice retreat from the goal of numerically solving the valuation function, giving up the seamless connection between economic theory and statistical inference for a parsimonious approximation of the value function. In principle both methods briefly reviewed here could be made as accurate as desired (choosing a sufficiently high dimensional subspace, only exchanging the integration and maximization operators at a sufficiently small number of nodes). More generally a secondary use of any reasonable approximating rule in an infinite horizon problem is to provide a useful initial starting point, to which value function contractions and/or policy iterations can be applied.

Rather than solve for the optimal decision rules, one could assume agents themselves choose rules from a smaller supspace. For example, suppose agents pick index rules which are linear in the state space (Hotz and Miller, 1988). This gives the estimation problem a standard index formulation, and thus dodges the issue of solving for the value function. In the job matching example, one would replace (6.20) with:

$$d_t = \begin{cases} 0 & \text{if} \quad \theta_0 + \theta_1\gamma_t + \theta_2\alpha_t > 0 \\ 1 & \text{if} \quad \theta_0 + \theta_1\gamma_t + \theta_2\alpha_t \leq 0. \end{cases} \tag{6.41}$$

Since d_t^o is (stepwise) increasing in γ_t and decreasing in α_t, one would expect an estimate of θ_1 to be positive, and a negative estimate of θ_2. While this greatly simplifies the estimation problem, the seamless transition from theory to data of a fully structural model is lost. In addition the misspecification induced by using a linear functional form (or more generally a parsimoniously specified nonlinear function) rather than the nonlinear mapping derived from the theory, $v_0 - v_1(\gamma_t, \alpha_t)$, induces a misspecification error: how serious this is presumably depends on the specifics of the application.

Another approximation which has been used in stopping problems (such as retirement from the workforce) is to exchange the order of the expectation and maximization operators, and thus redefine the agent's problem (Stock and Wise, 1990). Continuing the job matching model, let v_0^* denote the value of beginning a new job when this new rule is substituted for the optimal one. Rather than computing the lifetime value of staying on the job one more period and behaving optimally, each period the agent simply compares the expected value of staying on the current job forever, $\gamma_t/(1 - \beta)$, with the value of quitting now, v_0^*. So instead of (6.20) she sets:

$$d_t^* = \begin{cases} 0 & \text{if} \quad \gamma_t > (1 - \beta)v_0^* \\ 1 & \text{if} \quad \gamma_t \leq (1 - \beta)v_0^*. \end{cases} \tag{6.42}$$

Notice this rule delivers a higher expected lifetime utility than taking the first job and never switching, because the agent is more likely to move when matches are bad than good. Thus $v_0^* > 0$. But comparing d_t^* with d_t^o using (6.20) and

(6.42), clearly $v_0^* < v_0$, so the value from switching jobs is less under d_t^*. However, $\gamma_t \leq v(\gamma_t, \delta_t)$ too; under d_t^* the agent does not recognize the option value of retaining the current match now with a view to possibly breaking it in the future. Finally note that (6.42) is a special case of (6.41) which imposes the restrictions that $\theta_0 = (\beta - 1)v_0^*$, $\theta_1 = 1$, and $\theta_2 = 0$.

Finally, researchers are experimenting with the simultaneous use of several approximation techniques to obtain almost optimal rules at substantially reduced costs (Wolpin, 1992; Keane and Wolpin, 1994a). Intelligently chosen approximate value functions that pertain to nodes several periods after the current decision is made, interpolating the value function from a relatively small number of points in the state space (rather than attempting to compute the optimal rule on the whole state space), exchanging the expectation and maximization operators at judiciously chosen nodes, are three examples of the tools available. The basic thrust of this endeavor is to complement computational power with human cunning when hunting for optimal decision rules at a relatively low level of generality. It seems most appealing when, as in stationary infinite horizon models, a bound on the utility loss from using an approximation, versus the optimal rule, is easily derived.

5 Conditional Choice Probabilities

The previous section discussed how much numerical work is required to compute ML and related estimators. Although one might have thought faster computers would eventually overcome this limitation, paradoxically, advances in computer technology may have exacerbated it, by also making available to empirical workers ever increasing data sets. While greater access to more detailed and larger data sets has been a boon to researchers, encouraging them to analyze richer frameworks, the growth in the size of the data relative to the speed at which computers can solve dynamic programming problems provides additional motivation for researchers to look for alternative methods of estimation to ML. This section surveys a class of estimators, conditional choice probability (CCP) estimators, which have recently been developed to exploit larger data sets and economize on computational effort.

CCP estimators have been used to analyze fertility (Hotz and Miller, 1993), engine replacement (Hotz et al., 1994), stock replenishment (Aguirregabria, 1994), price changes (Slade, 1994), welfare participation (Miller and Sanders, 1996), and learning by doing (Altug and Miller, 1997). The intuition behind CCP estimation is that for a broad class of dynamic models of discrete choice, the probability of the econometrician observing a certain choice given the information in the data is closely related to the value of picking that choice. Rather than solving the dynamic programming problem directly, this approach uses sample frequencies to estimate the choice probabilities and obtains the expected value of making different choices as a mapping of those frequencies. Estimating

the structural parameters of interest then becomes the second part of a standard nonlinear two-stage estimation problem. A key feature of the CCP approach is the representation of the conditional value function as computationally tractable mapping of the choice probabilities. The basic idea is developed in two parts, first with a simple example and then with a more general result. This leads into a discussion of how the representation is applied in different estimation contexts and the resulting statistical properties.

5.1 Another Example

To illustrate the CCP approach, consider the following stopping problem. A publisher has given a textbook writer a deadline T. To simplify matters we will assume that if the deadline is not met, the project will lose all its value. Otherwise the current utility from submitting the text at time t is t^α. Therefore before T if the writer continues to work on the book he increases its value by one extra unit, but against this gain is the delay, that is discounting royalties by β per period. If there were no costs and benefits to consider, then the net marginal benefit of continuing is $\beta^t t^{\alpha-1}(\alpha + t \ln \beta)$, which has a unique root at $|\alpha/\ln \beta|$, and is positive for all t less than this crossing point. In this deterministic world the author continues writing in period $t \leq T$ if $t \leq |\alpha/\ln \beta| - 1$ and certainly submits his manuscript once $t > |\alpha/\ln \beta|$. Finally if $t < |\alpha/\ln \beta| < t + 1$ he submits at t or $t + 1$ depending on whether $\beta^t t^\alpha$ or $\beta^{(t+1)}(t + 1)^\alpha$ is higher.

Now suppose there are also some idiosyncratic factors associated with how easy it is to work in any given period, called ε_{0t}, and unanticipated time costs of dealing with his publisher, called ε_{1t}. It is further assumed that ε_{0t} and ε_{1t} are both independent and identically distributed random variables drawn from the type 1 extreme value probability distribution.[14] In terms of the notation developed in section 2, the state vector is $z_t = (\varepsilon_{0t}, \varepsilon_{1t})$, the utility received in periods before submitting the manuscript for publication is $u_{1t}(z_t) = \beta^t \varepsilon_{1t}$, while the utility from publishing on date t is $u_{0t}(z_t) = \beta^t(t^\alpha + \varepsilon_{0t})$. These random factors complicate the author's decision making because, other factors aside, he would prefer to submit when $\beta^t(\varepsilon_{0t} - \varepsilon_{1t})$ is high; there is a search aspect to consider.

The conditional value function for submitting his manuscript, denoted by setting $d_t = 0$, is $v_{0t} = \beta^t t^\alpha$, while the conditional value function for continuing to work at least one more period after t, or setting $d_t = 1$, is:

$$v_{1t} = E[v_{t+1}(\varepsilon_{0,t+1}, \varepsilon_{1,t+1})]. \tag{6.43}$$

In a similar vein, p_{kt}, the conditional choice probabilities for d_t, are defined as $p_{kt} = E[d_t = k \mid t]$ for the two actions $k \in \{0, 1\}$. Given the distributional assumptions, the conditional choice probability of choosing to extend the gestation phase one more period is $[1 + \exp(v_{1t} - v_{0t})]^{-1}$, while the analogous expression for p_{1t} simply exchanges the subscripts in this expression.[15] Taking the quotient and then the natural logarithm yields:

$$\ln p_{1t} - \ln p_{0t} = v_{1t} - v_{0t}. \tag{6.44}$$

This inversion result, generalized below, represents differences in the conditional value functions as a mapping from the conditional choice probabilities. From (6.44) it is easy to see that p_{1t} exceeds half if and only if v_{1t} exceeds v_{0t}. This is because $v_{1t} - v_{0t}$ represents the difference between the valuations of the options without taking account of the unobserved symmetrically distributed random variable $\varepsilon_{1t} - \varepsilon_{0t}$.

Some further algebra establishes that the expectation of ε_{kt} conditional on action $k \in \{0, 1\}$ also has a simple form:

$$E[\varepsilon_{kt} \mid d_t = k] = \zeta - \ln p_{kt}, \tag{6.45}$$

where $\zeta \cong 0.577$ is Euler's constant. The qualitative properties of this dynamic selection correction term are quite intuitive. From (6.44) the higher the difference $v_{1t} - v_{0t}$, the larger is p_{1t}, and therefore the threshold value of $\varepsilon_{1t} - \varepsilon_{0t}$ to induce the writer to continue working is lower, the expectation converging to its unconditional counterpart of ζ at $p_{1t} = 1$. This implies the conditional expectation of ε_{kt} is declining in p_{kt}. The conditional value function for working on the manuscript another period is thus:

$$\begin{aligned} v_{1t} &= p_{0,t+1}\{E[\varepsilon_{0t} \mid d_{0t} = 0] + v_{0,t+1}\} + p_{1,t+1}\{E[\varepsilon_{1t} \mid d_t = 1] + v_{1,t+1}\} \\ &= p_{0,t+1}\beta^{t+1}(t+1)^{\alpha} + p_{1,t+1}v_{1,t+1} + \beta^{t+1}\left(\zeta - \sum_{k=0}^{1} p_{k,t+1} \ln p_{k,t+1}\right) \\ &= \beta^{t+1}(t+1)^{\alpha} + p_{1,t+1}(\ln p_{1,t+1} - \ln p_{0,t+1}) + \beta^{t+1}\left(\zeta - \sum_{k=0}^{1} p_{k,t+1} \ln p_{k,t+1}\right). \end{aligned} \tag{6.46}$$

The second line uses forwards induction to exploit the new representation by substituting in equation (6.44). Interpreting the bottom line of equation (6.46), the first expression, $\beta^{t+1}(t+1)^{\alpha}$, is the present value of submitting in period $t+1$, net of the unobserved component. The final expression, $(\zeta - \Sigma_{k=0}^{1} p_{k,t+1} \ln p_{k,t+1})$, captures the expected value of the unobservables. The middle term, $p_{1,t+1} - \ln p_{0,t+1})$, is just $v_{1,t+1} - v_{0,t+1}$, but weighted to account for the possibility that with probability $p_{1,t+1}$ it is optimal to continue writing for more than one period longer.

There are two structural parameters of interest, the subjective discount rate β and the concavity parameter α. The econometric framework combines (6.43), (6.44), and (6.46) to obtain:

$$\ln \frac{p_{1t}}{p_{0t}} = \beta^{t+1}(t+1)^{\alpha} + p_{1,t+1} \ln \frac{p_{1,t+1}}{p_{0,t+1}} + \beta^{t+1}\left(\zeta - \sum_{k=0}^{1} p_{k,t+1} \ln p_{k,t+1}\right) - \beta^t t^{\alpha}. \tag{6.47}$$

The details of estimation are deferred to later: suffice it to remark here that equation (6.47) links together observed variables, that is t, and easily estimated

probabilities $(p_{0t}, p_{0,t+1})$ with the parameters characterizing the structural framework $\theta \equiv (\alpha, \beta)$.

5.2 A Representation Theorem

The illustration can be extended to a large class of dynamic programming problems. Returning to the discrete choice problem considered in section 2, we now impose an additional assumption which restricts the role of unobserved variables in the analysis:

$$\Pr\{x_{t+1}, \varepsilon_{t+1} \mid x_t, \varepsilon_t, d_t\} = \Pr\{\varepsilon_{t+1} \mid x_{t+1}\}\Pr\{x_{t+1} \mid x_t, d_t\}. \tag{6.48}$$

This assumption, introduced to the ML literature by Rust (1987) and widely used, is also key to the CCP estimation strategy. There are essentially two parts to it. Conditional on x_t, the unobserved variables ε_t are independent of their past values:

$$\Pr\{\varepsilon_{t+1} \mid x_{t+1}, x_t, \varepsilon_t, d_t\} = \Pr\{\varepsilon_{t+1} \mid x_{t+1}\} \tag{6.49}$$

The second part of the assumption restricts the law of motion of the observed variables. The only way the unobserved variable might affect outcomes is through the choices agents make:

$$\Pr\{x_{t+1} \mid x_t, \varepsilon_t, d_t\} = \Pr\{x_{t+1} \mid x_t, d_t\}. \tag{6.50}$$

Admittedly, assumption (6.48) is quite restrictive. For example it is violated in the job matching model discussed in section 3, because one of the state unobserved variables, subjective beliefs about the mean of the distribution for the job match, is autocorrelated.

As the publishing example demonstrated, the conditional choice probabilities play an important role because of their relationship to the conditional value functions. In this more general setting, the conditional value function for action $k \in \{0, \ldots, K\}$ is defined as:

$$v_{kt}(x_t) = u_{kt}(x_t) + E[v_t(z_{t+1}) \mid x_t, d_{kt} = 1], \tag{6.51}$$

and the conditional choice probability is:

$$p_{kt}(x_t) = \Pr\{d_{kt} = 1 \mid x_t\} = \Pr\left\{\bigcap_{j=1}^{K}[\varepsilon_{jt} \le \varepsilon_{kt} + v_{kt}(x_t) - v_{jt}(x_t)] \mid x_t\right\}. \tag{6.52}$$

Drop the t subscript from x_t, writing:

$$p_t \equiv p_t(x) \equiv (p_{1t}(x), \ldots, p_{Jt}(x))' \equiv (p_{1t}, \ldots, p_{Jt})' \tag{6.53}$$

for the vector of conditional choice probabilities, and let $\psi_t \equiv \psi_t(x)$ denote the difference in the conditional value functions from any benchmark value function, say $v_{0t}(x)$, so that:

$$\psi_{jt} \equiv \psi_{jt}(x) \equiv v_{jt}(x) - v_{0t}(x). \tag{6.54}$$

Equation (6.52) may now be expressed as:

$$p_{kt} = \int_{-\infty}^{\infty} G_{kt}(\ldots, \varepsilon_{kt} + \psi_{kt} - \psi_{k-1,t}, \varepsilon_{kt}, \varepsilon_{kt} + \psi_{kt} - \psi_{k+1,t}, \ldots \mid x) d\varepsilon_{kt}, \quad (6.55)$$

where $G_{kt}(\varepsilon_{0t}, \ldots, \varepsilon_{Kt} \mid x) \equiv \partial G_t(\varepsilon_{0t}, \ldots, \varepsilon_{Kt} \mid x)/\partial \varepsilon_{kt}$, is the derivative with respect to ε_{kt} of $G_t(\varepsilon_{0t}, \ldots, \varepsilon_{Kt} \mid x_t)$, the joint probability distribution function for the unobserved variables $(\varepsilon_{0t}, \ldots, \varepsilon_{Kt})$ conditional on the observed variables x_t. Given any x, suppose for a moment that the K-dimensional vector p_t is known, but that the K-dimensional vector ψ_t is unknown. Then ψ_t could be solved at x if a unique solution existed for the K equations (6.55). Equation (6.44) proves, by construction, the existence of a unique solution in the publishing example described above. More generally, a representation theorem proved in Hotz and Miller (1993) establishes (6.55) is indeed invertible in ψ_t: for each x there exists a unique vector ψ_t, which solves the K equations.

This theorem also paves the way for expressing the expected values of the unobserved variables, which are drawn from a sample that is subject to choice based censoring. To see this, let $Q_t(p_t, x)$ denote the inverse function of (6.55), implying $\psi_t(x) = Q_t(p_t, x)$. Then the expected value of ε_{kt} conditional on choice k and observed state x is:

$$\begin{aligned} w_{kt}(p_t, x) &\equiv E[\varepsilon_{kt} \mid x, d_{kt} = 1] \\ &= \int_{-\infty}^{\infty} \varepsilon_{kt} G_{kt}(\ldots, \varepsilon_{kt} + Q_{kt}(p_t, x) - Q_{k-1,t}(p_t, x), \varepsilon_{kt}, \varepsilon_{kt} \\ &\quad + Q_{kt}(p_t, x) - Q_{k+1,t}(p_t, x), \ldots \mid x) d\varepsilon_{kt}. \end{aligned} \quad (6.56)$$

Knowing he will behave optimally, the utility an agent expects to receive in a future period if the observed part of the state x occurs is therefore:

$$\sum_{k=0}^{K} p_{kt}(x)[u_{kt}(x) + w_{kt}(p_t, x)]. \quad (6.57)$$

5.3 Applying the Representation Theorem

In principle one could sum probability weighted utilities obtained from (6.57) over periods and states to obtain an approximation for the value function. Using this procedure in estimation would yield considerable savings in computation time relative to ML, because the sample probabilities could be directly used as weights, rather than having to solve for the optimal decision rule as a function of the underlying parameters and the state variables. However when the dynamic programming problem has a special structure, it can be further exploited to reduce the computational burden. To date three classes of problems have been discovered: problems with terminating actions, problems with finite dependence, and stationary Markov problems.

The first special case we examine occurs when there exists an action, called a terminating action, that ends the decision making phase (Hotz and Miller, 1993). Suppose action zero is a terminating action (as in the publishing example). Having taken the terminating action once, the agent is compelled to live with repeating that action for the rest of his life. In this case:

$$v_{0t}(x) = u_{0t}(x) + E\left[\sum_{s=t+1}^{T} u_{0s}(x_s) + \varepsilon_{0s} \mid x\right], \quad (6.58)$$

which may be calculated directly (as a function of some structural parameters to be estimated). Therefore, by the representation theorem all the other value functions can be calculated as:

$$v_{jt}(x) = Q_{jt}(x) + u_{0t}(x) + E\left[\sum_{s=t+1}^{T} u_{0s}(x_s) + \varepsilon_{0s} \mid x\right]. \quad (6.59)$$

Another case which further reduces the computational demands of estimation occurs when the state variables only exhibit finite time dependence (Altug and Miller, 1997). To exploit the finite dependence property in applying the representation theorem, the following notation is handy. First the transpose of the state, $z_t' = (x_t', \varepsilon_t')$, is partitioned into previous choices that enter it, $(d_{t-1}', \ldots, d_{t-\rho}')$, a vector of observed, strictly exogenous variables denoted y_t', and the unobserved variables entering the state ε_t. Thus $x_t' \equiv (y_t', d_{t-1}', \ldots, d_{t-\rho}')$, the integer ρ indicating the length of dependence. Imagine calculating $v_{kt}(x_t)$ in the following manner. The agent sums up the utility from taking action k this period, and action 0 in the next ρ periods, behaving optimally from then on, and finally adding a correction factor to account for her nonoptimal actions in periods $t + 1$ through $t + \rho$. Accordingly, let $z_{k,t+s}^{t}$ denote the state variable in period $t + s$ associated with this action sequence, and $x_{k,t+s}^{t}$ its observed part. By inspection $z_{k,t+\rho+1}^{\rho+1} = (y_{t+\rho+1}', 0, \ldots, 0, \varepsilon_{t+\rho+1}')$ is strictly exogenous, and therefore does not depend on k. This action sequence and subsequent behavior generates a remaining lifetime utility of $v_{t+\rho+1}(z_{t+\rho+1}^{\rho+1})$ from period $t + p + 1$, and the utility received during a typical period $s \in [t, \ldots t + \rho]$ in the meantime is $u_{0s}(x_{k,t+s}^{t}) + \varepsilon_{0s}$. Finally the expected loss, based on information at period t from taking action 0 in period $s \in [t + 1, \ldots t + \rho]$ instead of the optimal one, is:

$$E\left\{\sum_{s=t+1}^{\rho+t} \sum_{j=0}^{K} p_{js}(x_{ks}^{s-t})[Q_{js}(p_s(x_{ks}^{s-t}), x_{ks}^{s-t}) + w_{js}(p_s(x_{ks}^{s-t}), x_{ks}^{s-t})] \mid x_t\right\}. \quad (6.60)$$

The first group of expressions in (6.60), comprising the $p_{js}(x_{ks}^{s-t})Q_{js}(p_s(x_{ks}^{s-t}), x_{ks}^{s-t})$ terms, is attributable to taking the benchmark action (rather than the optimal one at s) as they affect utility through the observed state variables, while the terms like $p_{js}(x_{ks}^{s-t})w_{js}(p_s(x_{ks}^{s-t}), x_{ks}^{s-t})$ affect utility through the expected value of unobserved variables. Combining the various terms we thus obtain:

$$v_{kt}(x_t) = u_{kt}(x_t) + E[v_{t+\rho+1}(z_{k,t+\rho+1}^{\rho+1}) \mid x_t]$$
$$+ E\left\{ \sum_{s=t+1}^{\rho+t} u_{0s}(x_{kt}^{s-t}) + \sum_{j=0}^{K} p_{js}(x_{kt}^{s-t})[Q_{js}(p_s(x_{kt}^{s-t})) + w_{js}(p_s(x_{kt}^{s-t}), x_{kt}^{s-t})] \mid x_t \right\}. \tag{6.61}$$

Stationary Markov problems whose observed characteristics come from a finite set also have a special structure which can be exploited in estimation to further reduce the computational burden (Aguirregabria, 1994; Hotz et al., 1994; Rust, 1995). Assuming current utility is discounted at a geometric rate over time, $u_{kt}(z)$ simplifies to $\beta^t[u_k(x) + \varepsilon_{jt}]$, and (6.5) reduces to:

$$E\left[\sum_{t=0}^{\infty} \sum_{k=0}^{K} d_{kt}\beta^t[u_k(x) + \varepsilon_{jt}] \mid z_0\right]. \tag{6.62}$$

Defining the observed part of the state space by the set $\{x^{(1)}, \ldots, x^{(L)}\}$ for some positive integer L, the objects of the problem can be expressed more concisely by writing the observed parts of current utility as $u_j^{(l)}$ rather than $u_j(x^{(l)})$, the unobserved parts as $w_j^{(l)}$ instead of $w_j(p_j(x^{(l)}), x^{(l)})$, the conditional value functions as $v_j^{(l)}$ not $v_j(x^{(l)})$, the conditional probabilities as $p_j^{(l)}$ not $p(x^{(l)})$, and the conditional transition probabilities as $F_j^{(ll')}$ rather than $F_j(x^{(l')} \mid x^{(l)})$. Therefore the probability of transiting from $x^{(l)}$ to $x^{(l')}$, denoted by $T_{ll'}$, is $T_{ll'} = \Sigma_{j=0}^{J} p_j^{(l)} F_j^{(ll')}$ and the transition matrix for the observed states is thus defined as:

$$T = \begin{bmatrix} T_{11} & \cdots & T_{1L} \\ \vdots & \ddots & \vdots \\ T_{L1} & \cdots & T_{LL} \end{bmatrix}. \tag{6.63}$$

Let $\underline{u}$ denote the $L \times 1$ vector of expected utilities (that is conditional on $x^{(l)}$ but margining over the unobserved variables and the choices):

$$\underline{u} = \begin{bmatrix} \Sigma_{j=0}^{J} p_j^{(1)}(u_j^{(1)} + w_j^{(1)}) \\ \Sigma_{j=0}^{J} p_j^{(L)}(u_j^{(L)} + w_j^{(L)}) \end{bmatrix}. \tag{6.64}$$

Defining the L-dimensional row vector of conditional probability transitions as $F_j^{(l)} \equiv (F_j^{(l1)}, \ldots, F_j^{(lL)})$, the conditional value functions may be expressed in the new notation as:

$$v_j^{(l)} = u_j^{(l)} + \beta F_j^{(l)} \sum_{s=0}^{\infty} \beta^s T^s \underline{u}, \tag{6.65}$$

where T^s is the s-step transition matrix found by multiplying T with itself s times. As a practical matter, approximating the infinite series with a large finite number is a useful way of approximating $v_j^{(l)}$ for moderate values of β. Alternatively, note that the infinite geometric matrix sum, $\Sigma_{s=0}^{\infty}\beta^s T^s$, is the inverse of $(I_L - \beta T)$, where I_L is the $L \times L$ identity matrix. So whenever the inverse of $(I_L - \beta T)$ is cheap to compute, (6.65) can be directly calculated from:

$$v_j^{(l)} = u_j^{(l)} + \beta F_j^{(l)}(I_L - \beta T)^{-1}\underline{u}. \tag{6.66}$$

5.4 Estimation

When the optimization problem exhibits one (or more) of these three properties – terminal states, finite dependence, or a stationary Markov nature – the conditional value functions can be easily expressed using formulas (6.59), (6.61), or (6.66). A two-stage procedure can then be used to estimate the structural parameters. Here we consider the case in which the observed portion of the state space is finite.[16] First estimates for the choice and transition probabilities are obtained from the relative frequencies observed in the sample. Then the structural parameters are estimated by optimizing a criterion function that uses the results from the first stage.

There are several related ways of constructing a criterion function to undertake the second stage. As before, the observed part of the state space is labeled $x^{(1)}$ through $x^{(L)}$. Accordingly, let $\theta \in \Theta$ denote the unknown structural parameters to be estimated, and let P denote the vector of conditional choice probabilities. In the finite horizon model P has dimension LJT because it is arrayed over the observed states, dates, and choices, while in the stationary Markov model its dimension is only LJ. We suppose the true value of the incidental and structural parameters are respectively P_0 and θ_0, and denote by P_N the cell estimators obtained from the first stage of estimation. Denote by $v_{kt}(x^{(l)}, \theta, P)$ the conditional value function, using the representation described above, for taking action k at date t when the observed part of the state space is $x^{(l)}$. To conduct the second stage of estimation, a quasi-log-likelihood function could be formed from:

$$\ell_n(\theta) = \ln \Pr\left\{ j = \underset{k \in \{0,\ldots,J\}}{\operatorname{argmax}} [\varepsilon_{nk} + v_{kt}(x_n, \theta, P_N)] \mid x_n \right\}, \tag{6.67}$$

where x_n comprise the observed characteristics of the nth observation which we suppose is sampled at t. When the structural parameter estimates have been computed, the standard errors are calculated to account for the two-stage estimation strategy, using the standard correction formula (Newey, 1984).

Minimum distance (MD) estimation provides another means of obtaining an $N^{1/2}$ consistent and asymptotically normal estimate of θ_0 (Chamberlain, 1984). Applied here, it is based directly on the equality $\psi_t(x) = Q_t(p_t, x_t)$. Accordingly define for each $\theta \in \Theta$ the LJT-dimensional vector $\psi(\theta, P_N)$ as:

$$\psi(\theta, P_N)' \equiv (\psi_1^{(1)}(\theta, P_N)', \ldots, \psi_1^{(L)}(\theta, P_N)', \psi_2^{(1)}(\theta, P_N)', \ldots, \psi_T^{(L)}(\theta, P_N)') \tag{6.68}$$

in the finite horizon model where:

$$\psi_t^{(l)}(\theta, P_N) \equiv (v_{1t}(x^{(l)}, \theta, P_N) - v_{0t}(x^{(l)}, \theta, P_N), \ldots, v_{Jt}(x^{(l)}, \theta, P_N) - v_{0t}(x^{(l)}, \theta, P_N)). \tag{6.69}$$

Similarly define $Q(\theta, P)$ as the LJT-dimensional vector in the finite (infinite) horizon model, its components formed from $Q_t(p_t(x^{(l)}), x^{(l)}, \theta)$ for each (t, l). (In the infinite horizon stationary case the dimension of $\psi(\theta, P_N)$ and $Q(\theta, P)$ is only LJ and the t subscripts are dropped in (6.68) and (6.69). An MD estimator can now be defined for θ_0 by minimizing:

$$N(Q(\theta, P_N) - \psi(\theta, P_N))' A_N (Q(\theta, P_N) - \psi(\theta, P_N)) \tag{6.70}$$

in $\theta \in \Theta$, where A_N is any LJ square matrix converging in probability to some positive definite weighting matrix A. Let $D_\theta \equiv \partial(Q(\theta_0, P_0) - \psi(\theta_0, P_0))/\partial\theta$ denote the partial derivative vector of $Q(\theta_0, P_0) - \psi(\theta_0, P_0)$ with respect to θ evaluated at the true conditional choice probabilities P_0 and the true values of the structural parameters θ_0, and define $D_P \equiv \partial(Q(\theta_0, P_0) - \psi(\theta_0, P_0))/\partial P$ similarly as the partial derivative vector for the conditional choice probabilities. Noting that $(\sqrt{N})(P_N - P_0)$ is asymptotically normal with mean 0 and variance V (which is block diagonal by state and date), it follows by taking a first-order Taylor series expansion of the FOC for (6.69) that $(\sqrt{N})(\theta_N - \theta_0)$ is asymptotically normal with mean 0 and variance $(D_\theta A D_\theta)^{-1}(D_\theta A D_P)V(D_P A D_\theta')(D_\theta A D_\theta)^{-1}$. With regards to small sample properties, a Monte Carlo study undertaken by Hotz et al. (1994) suggests that the CCP estimator exhibits more finite sample bias than ML when the underlying structural assumptions are correctly specified, but the loss in asymptotic efficiency is not severe.

This section concludes by briefly revisiting the example which motivated it. Let τ_n denote the amount of time the nth author spends writing his textbook, and suppose some publishing firms have records on N textbooks, the data set comprising a finite sequence $\{\tau_n\}_{n=1}^{N}$. Empirically implementing this model involves estimating the particular environment which generated these data, namely $\theta_0 \equiv (\alpha_0, \beta_0)$. For expositional simplicity, suppose there is no unobserved hetereogeneity apart from the transitory disturbances $(\varepsilon_{0t}, \varepsilon_{1t})$. Thus $x \equiv t$. Although the conditional choice probabilities are unknown, they can be easily estimated from their sample frequencies. Accordingly we define $P_N = (p_{01}^{(N)}, \ldots, p_{0T}^{(N)})'$ as:

$$p_{0t}^{(N)} = \sum_{n=1}^{N} 1(\tau_n = t) / \sum_{n=1}^{N} 1(\tau_n \geq t), \tag{6.71}$$

where $1(A)$ is the indicator function for the statement A. Appealing to (6.47), we now form the vector $\psi(\theta, P_N) \equiv (\psi_1(\theta, P_N), \ldots, \psi_T(\theta, P_N))$ using the cell estimates, with a generic component defined by the expression:

$$\psi_t(\theta, P_N) = \beta^{t+1}(t+1)^\alpha + p_{1,t+1}^{(N)} \ln \frac{p_{1,t+1}^{(N)}}{p_{0,t+1}^{(N)}} + \beta^{t+1}\left(\zeta - \sum_{k=0}^{1} p_{k,t+1}^{(N)} \ln p_{k,t+1}^{(N)}\right) - \beta^t t^\alpha, \tag{6.72}$$

where $p_{1t}^{(N)} \equiv 1 - p_{0t}^{(N)}$, and proceed with the MD estimation strategy described above.

6 Time Additive Continuous Choice Models

This section and the following one discuss the estimation of structural models in which individuals from the sample make continuous choices. Much recent work in this area estimates parameters from the first-order condition (FOC) to an optimization problem which is solved within an equilibrium for the population. Because the FOC is an equality which holds conditional on previous actions and the current state, orthogonality conditions may be constructed from multiplying the FOC with predetermined and exogenous variables. If, in addition, the unobserved parts of the first-order condition are uncorrelated with these predetermined variables and have mean zero, then the product of the predetermined variables and the observed parts of the first-order condition also have mean zero in expectation at the true parameter values. Providing a sample average of realizations drawn from a cross-section or panel converges to this expectation, estimators with the standard large sample properties can be found.

The difficulties encountered in estimation revolve around picking the predetermined variables, or instruments, and interpreting the empirical results. These difficulties stem from three sources. Relative price movements are common shocks that affect all consumers the same way, and as elaborated below, affect the scope for choosing instruments, because a sample moment taken over a cross-section does not have the same asymptotic properties as a sample moment from time series observations (Chamberlain, 1984). Here the assumption of complete and competitive markets (CCM) can play a major role in resolving this issue, but restrictions implied by the weaker permanent income hypothesis (PIH) are insufficient to identify the model. Nonadditive preferences in observed variables constitute a second source of complications in estimation, even when there are no aggregate fluctuations. Serial correlation in unobserved variables that are specific to the individual also reduces the set of instruments available for research exploiting panel data (ruling out past choices for example): in some parameterizations this can be handled using the usual procedures of differencing logarithms and so forth, but when preferences are not separable over time this may be unfeasible.

6.1 Time Separable Preferences

If the agent does not make discrete choices, then equation (6.4) simplifies to:

$$E\left[\sum_{t=0}^{T} u_t(c_t, z_t) \mid z_0\right], \tag{6.73}$$

and the k subscript on the right side of equation (6.3), the transition probability distribution function for the state variable, can be dropped. Put succinctly, the assumption of time separability states that the state variable next period does not depend on the current choice. Thus, when the support for z is continuous:

$$\Pr\{z_{t+1} \leq z \mid z_t, c_t\} \equiv F_t(z \mid z_t). \tag{6.74}$$

This does not mean that choices are uncorrelated across periods. For example, if z_t is a time dependent process, and $u_t(c_t, z_t)$ is nonseparable in its two arguments, then optimal behavior induces serial correlation in c_t. Nevertheless an inspection of (6.73) and (6.74) reveals that the expected lifetime utility gain from consuming an extra unit now is calculated without any reference to the future. Therefore only the cost of marginal consumption can trigger dynamics if preferences are separable over time.

6.2 The Lifetime Budget Constraint

The most natural generalization of perfect foresight models of lifecycle behavior is the assumption that markets are competitive and complete (CCM). Here the word "competitive" is synonymous with price taking behavior; "complete markets" means there are no frictions in the markets for loans, a common interest rate facing borrowers and lenders, and that a rich set of financial securities exists to hedge against uncertainty. The chief virtue of assuming CCM is that it incorporates uncertainty in a sufficiently simple way to yield a tractable econometric model, because under CCM aggregate effects are fully transmitted through prices to contingent claims (Altug and Miller, 1990).

To formally define the lifetime budget constraint the commodity and price spaces must be properly specified. In a standard general equilibrium model all trades take place at date 0, history only determining the realized path of deliveries and consumptions. Accordingly let $\Lambda_t(z)$ denote the price measure for the first good when the date/state coordinate pair is (t, z) and denote by $\lambda_t(z)$ its (Radon–Nikodym) derivative. In other words, to consume a unit of this *numéraire* good in state $z \in A$ at date t costs:

$$\int_{z \in A} \lambda_t(z) \mathrm{d}F_t(z \mid z_0), \tag{6.75}$$

where $F_t(z \mid z_0)$ is the probability distribution function for Z at date t conditional only on the initial state z_0 at date 0, formed from successive convolutions to (6.74). Also let $\varphi(z) = (1, \varphi_2(z), \ldots, \varphi_M(z))$ denote an M-dimensional row vector of spot prices in each state z, defined in terms of the first good. Under CCM the household must only obey a lifetime budget constraint of the form:

$$w \geq E\left[\sum_{t=0}^{t=T} \lambda_t \varphi_t c_t \mid z_0\right], \tag{6.76}$$

where w is lifetime wealth. For example if there are only L (finite) states (6.76) would simplify to:

$$E\left[\sum_{t=0}^{t=T} \lambda_t \varphi_t c_t \mid z_0\right] \equiv \sum_{t=0}^{t=T} \sum_{l=1}^{L} \left\{\lambda_{lt} [F_t(l \mid z_0) - F_t(l-1 \mid z_0)] \sum_{m=1}^{M} \varphi_{lmt} c_{lmt}\right\}, \tag{6.77}$$

where $\lambda_{lt}\varphi_{lmt}[F_t(l \mid z_0) - F_t(l-1 \mid z_0)]$ is the contingent price of consuming a unit of good m on date t if state l occurs, that is the product of the spot price φ_{lmt} of the good at t, and the contingent price of the *numéraire* good, $\lambda_{lt}[F_t(l \mid z_0) - F_t(l-1 \mid z_0)]$.

Equation (6.76) expresses the budget constraint in an analogous way to a budget constraint for a static problem (Debreu, 1959). It can also be recast in recursive fashion (Arrow, 1963). Rather than defining all state contingent claims on each date for the *numéraire* good, contingent claims markets are only open for the next period (and thus depend on the state of the current period). In the discrete state model this reduces the number of contingent claims markets from L^T to LT. Concerning the price of the one-period security, given (t, z), let $\Lambda_t(z^* \mid z)$ be the conditional price measure whose Radon–Nikodyn derivative $\lambda_t(z^* \mid z)$ satisfies:

$$\int_{z \in A} \left\{ \int_{z^* \in A} \lambda_t(z^* \mid z) \mathrm{d}F_t(z^* \mid z) \right\} \lambda_t(z) \mathrm{d}F_t(z \mid z_0) = \int_{z^* \in A} \lambda_{t+1}(z^*) \mathrm{d}F_{t+1}(z^* \mid z_0) \tag{6.78}$$

for all $A \subseteq Z$. Intuitively (6.78) states that the date 0 price of consuming a unit of the *numéraire* in states A on date $t+1$ (the right side of the equation) equals the date 0 cost of $[\int_{z^* \in A} \lambda_t(z^* \mid z) \mathrm{d}F_t(z^* \mid z)]$ securities that have payoffs in each of the states $z \subseteq Z$ at period t, which is then used to finance a unit of the *numéraire* in date $t+1$ to be received if $A \subseteq Z$ occurs. The agent's problem can now be written as choosing $(c_t, w_t(z^*)) \equiv (c_t(z), w_t(z^* \mid z))$ to maximize:

$$u_t(c_t, z_t) + \int v_t(z^*, w^*) \mathrm{d}F(z^* \mid z_t) \tag{6.79}$$

subject to the constraint that:

$$w_t - E[\varphi_t c_t + \lambda_t(z^* \mid z_t) w_t(z^* \mid z_t) \mid z_t] \geq 0 \tag{6.80}$$

and the law of motion for w_t, which is simply $w_{t+1} = w_t(z_{t+1} \mid z_t)$. (In the discrete case there are just M goods to pick for immediate consumption and L assets to choose from.) Recalling the framework of section 2, the state space for this reformulation of the problem requires w_t to be one for the components in the z vector. With reference to (6.2) there is only one index function, namely the left side of (6.80). Therefore the consumer's problem of maximizing (6.73) subject to (6.76) also fits within the framework that section 2 described.

6.3 Estimation and Testing CCM

To develop some intuition for the restrictions implied by CCM, we first consider the spot market for two goods, say the first two in the consumption vector. Assuming an interior solution, and first differencing the FOC obtained by maximizing (6.73) subject to (6.76), we obtain:

$$\ln\{\partial u_t(c_{nt}, z_{nt})/\partial c_{1nt}\} - \ln\{\partial u_t(c_{nt}, z_{nt})/\partial c_{2nt}\} = \ln\{\lambda_{1t}\} - \ln\{\lambda_{2t}\} \tag{6.81}$$

after taking the logarithm, where the n subscript now designates one of the

N people sampled in the data. Equation (6.81) states that, in equilibrium, the (logarithm of the) marginal rate of substitution between two goods consumed simultaneously equals the (logarithm) of their relative prices. Analogous to the notational convention developed in section 3, let $z_{nt} = (x_{nt}, \varepsilon_{nt})$, where x_{nt} is an observed variable but ε_{nt} is not, and suppose:

$$\ln [\partial u_t(c_{nt}, z_{nt})/\partial c_{1nt}] - \ln [\partial u_t(c_{nt}, z_{nt})/\partial c_{2nt}] = m_{12t}(c_{nt}, x_{nt}, \theta_0) + \varepsilon_{12nt}, \quad (6.82)$$

where $m_{12t}(c_{nt}, x_{nt}, \theta_0)$ is a function of observed variables (c_{nt}, x_{nt}), known up to a parameter vector $\theta_0 \in \Theta$ to be estimated, and ε_{12nt} is formed from ε_{nt}. If there are instruments y_{nt} satisfying the condition:

$$E[\varepsilon_{12nt} y_{nt}] = 0 \quad (6.83)$$

then conditions (6.82) and (6.83) can be exploited in estimation because together they imply:

$$0 = \operatorname*{plim}_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} y_{nt}[m_{12t}(c_{nt}, x_{nt}, \theta_0) - \ln [\lambda_{1t}] + \ln [\lambda_{2t}]]. \quad (6.84)$$

Stack $y_{nt}[m_{12t}(c_{nt}, x_{nt}, \theta) - \ln [\lambda_{1t}] + \ln [\lambda_{2t}]]$ over the periods covered by the panel to obtain the vector:

$$f_n(\theta) = \begin{bmatrix} y_{n1}[m_{121}(c_{n1}, x_{n1}, \theta) - \ln [\lambda_{11}] + \ln [\lambda_{21}]] \\ \cdots \\ \cdots \\ y_{nT}[m_{12T}(c_{nT}, x_{nT}, \theta) - \ln [\lambda_{1T}] + \ln [\lambda_{1T}]] \end{bmatrix}. \quad (6.85)$$

A generalized method of moments (GMM) estimator may now be defined by taking a quadratic form of the sample average of $f_n(\theta)$, and minimizing it with respect to $\theta \in \Theta$ (Hansen, 1982). Accordingly define θ^N as:

$$\theta^N = \operatorname{argmin}\left\{N\left[\sum_{n=1}^{N} \frac{f_n(\theta)'}{N}\right] W_N \left[\sum_{n=1}^{N} \frac{f_n(\theta)}{N}\right]\right\}. \quad (6.86)$$

where W_N is a positive definite weighting matrix. Typically a standard two-round estimation procedure is applied. In the first round the identity matrix $W_N = I$ is used as the weighting matrix to obtain consistent parameter estimates θ_N, and in the second any consistent estimator of the inverse of $\Sigma = E[f_n(\theta_0) f_n(\theta_0)']$. The resulting estimator is $N^{1/2}$ consistent and asymptotically normal with covariance matrix $(D'\Sigma^{-1}D)^{-1}$, where $D = E[\partial f_n(\theta_0)/\partial\theta]$. Moreover if the dimension of $f_n(\theta)$ is greater than the dimension of θ, then, under the null hypothesis that the model is correctly specified, the minimized criterion function (6.86) converges to a chi-square random variable with degrees of freedom equal to the number of overidentifying restrictions (Gallant and Jorgenson, 1979; Hansen, 1982).

Having seen how to estimate the marginal rate of substitution between two goods consumed at the same time, the extension to the case of the same good consumed in successive periods is straightforward. Following (6.81), let Δ denote the first difference operator and note:

$$\Delta \ln [\partial u_t(c_{nt}, z_{nt})/\partial c_{1nt}] = \ln [\lambda_{1,t+1}] - \ln [\lambda_{1t}]. \quad (6.87)$$

That is, the logarithm of the marginal rate of substitution between consumption in two consecutive periods equals the (logarithm) of their relative prices. Following the spirit of (6.82), write:

$$\Delta \ln [\partial u_t(c_{nt}, z_{nt})/\partial c_{1nt}] = m_{1t}(c_{n,t+1}, c_{nt}, x_{n,t+1}, x_{nt}, \theta_0) + \varepsilon_{1nt}, \quad (6.88)$$

where $m_{1t}(c_{nt}, x_{nt}, \theta_0)$ is a function of observed variables (c_{nt}, x_{nt}), known up to a parameter vector $\theta_0 \in \Theta$ to be estimated. If there are instruments y_{nt} satisfying the equality $E[\varepsilon_{1nt} y_{nt}] = 0$, then it can be exploited in estimation in practically the same way as above. There are two differences. When comparing two goods consumed in the same period, spot prices may be available in the data (or in the case of the labor market, a measure of the real wage), whereas comparing consumption of the same good in successive periods, the state contingent price is never observed. (Even in a world of perfect foresight it is not clear how the real interest rate should be measured.) Whereas values for $\ln [\lambda_{t1}] - \ln [\lambda_{t2}]$, the ratio of spot prices, could be inserted in (6.85), time dummies (indicator variables for the period) should be used to estimate the intertemporal condition. The second difference is attributable to the time additive parameterization itself. Since $m_{1t}(c_{n,t-1}, c_{nt}, x_{n,t+1}, x_{nt}, \theta_0)$ comes from $\partial u_{t+1}(c_{n,t+1}, z_{n,t+1})/\partial c_{1n,t+1}$ and $\partial u_t(c_{nt}, z_{nt})/\partial c_{1nt}$, which have the same functional form, whereas $m_{12}(c_{nt}, x_{nt}, \theta_0)$ is related to $\partial u_t(c_{nt}, z_{nt})/\partial c_{1nt}$ and $\partial u_t(c_{nt}, z_{nt})/\partial c_{2nt}$, which clearly have different functional forms, more parameters could be identified in one or the other formulation, depending on exactly how $u_t(c_{nt}, z_{nt})$ is specified, or the underlying identifying assumptions. In particular identifying assumptions stating the contemporaneous nonseparabilities between the components of the consumption vector and how the x_{nt} enter the respective marginal utility functions determine which combinations of parameters can be estimated.

While the assumption of competition between consumers seems innocuous, the notion that markets are complete is much harder to accept at face value. For example it implies that actuarially fair rates insure all losses to individuals that are independently distributed across a large population. But appearances can be deceptive. First, the aggregation results of Rubinstein (1981) show that, apart from personal insurance against idiosyncratic risk, only a small number of markets are required to exhaust trading opportunities: in a deterministic world just a one-period interest rate plus spot markets. For the utility parameterizations that most researchers focus on (the HARA class), one extra security suffices. CCM does not add that many restrictions to the data given the parametric assumptions that researchers are already making in the other parts of their framework. Moreover tests of CCM are primarily about the existence of spot markets and resource allocation; lacking data on food prices, for example, these tests do not distinguish between an allocation mechanism due to markets versus a food stamp program. All that empirical workers identify is whether people equate their marginal rate of substitution between pairs of commodities at their respective consumption points.

Apart from testing the overidentifying restrictions, a second way of testing the CCM null hypothesis is to add variables that under the null have no systematic effect. The alternative hypothesis is usually characterized by the statement that consumption is excessively volatile. To implement the test let h_{nt} be a vector of observed variables which do not enter the marginal rate of substitution function and replace (6.87) with:

$$\Delta \ln [\partial u_t(c_{nt}, z_{nt})/\partial c_{1nt}] = \ln \{\lambda_{1,t+1}\} - \ln \{\lambda_{1t}\} + \alpha h_{nt}. \tag{6.89}$$

Conducting the same estimation procedure as before, it follows that the null hypothesis would be rejected if the estimate of α differs significantly from 0. This approach is informative, but should be seen in context. If h_{nt} is correlated with unobserved variables in the utility function, the estimated coefficient would be nonzero even if the parameterization is correctly specified and CCM holds. Tests which ignore such correlations run the risk of falsely rejecting the null hypothesis.

Several studies following Altug and Miller (1990) have referred to the CCM assumption as a model of full insurance.[17] Focusing narrowly on the insurance aspects disguises the many other reasons why the framework might be rejected. None of the tests of full insurance maintain the null hypothesis of CCM in a vacuum, but simultaneously make assumptions about the separability of commodities, measurement error, the distribution of the unobservables, stationarity of the series, and of course the parameterization itself. Thus it is only possible to reject all of the maintained hypotheses at once, or alternatively, characterize the assumptions under which CCM is not rejected. Worse yet, testing CCM can be undertaken in a perfect foresight economy, and there is nothing intrinsic to these frameworks to suggest that uncertainty is playing any role at all. In a perfect foresight model consumers simply maximize:

$$\sum_{t=0}^{T} u_t(c_t, z_t) \tag{6.90}$$

subject to:

$$\sum_{t=0}^{t=T} \lambda_t c_t \leq w \tag{6.91}$$

and a known sequence $\{z_t\}_{t=1}^{T}$. If utility is additive over time, exactly the same set of first-order conditions (and estimation equations) emerge. Attributing model misspecification to a breakdown in insurance arrangements, other valid reasons why the framework might be misspecified (such as nonseparable preferences, poorly defined and/or poorly enforced property rights, and so on) are glossed over.

7 Relaxing Time Additivity in Continuous Choice Models

There have been several attempts to relax the assumption of time additivity (CCM and separable preferences), and they may be put into one of three categories. The most straightforward approach is to note that this assumption can be tested within and between baskets of goods and subpopulations: for example, consumption smoothing across generations but within dynasties (Hayashi, Altonji, and Kotlikoff, 1996), between inhabitants of the same Indian village versus people belonging to different villages (Townsend, 1994), and within different regions of a country versus across currency and national borders (Miller and Sieg, 1997). This approach is discussed first.

It is more ambitious to estimate preference parameters without making any assumptions about the trading opportunities agents face, by relying only on the properties of time additive preferences and rationality. Models of the permanent income hypothesis (PIH), the name under which this approach goes, have an analog in the time series literature, where national consumption aggregates have been modeled as giving utility to a representative consumer (Hansen and Singleton, 1982). In contrast to the macroeconometrics literature, which appeals to the asymptotic behavior of stationary time series to undertake estimation and hypothesis testing, empirical studies of the PIH using micro data rely on cross-sectional averages to conduct statistical inference. However the fluctuations in aggregate time series, which are after all the very basis for empirical work in macroeconometrics, are, simultaneously, incontrovertible evidence that individuals within the population experience common shocks. Whereas we have seen that CCM provides investigators with an analytic tool for dealing with these common shocks in an internally consistent manner through the insertion of time dummies, the PIH yields insufficient structure to achieve identification. Unfortunately this stumbling block has not prevented the growth of a literature which simply ignores these issues.

After critically reviewing the problems associated with estimating models of the PIH with micro data, this survey chapter concludes by reviewing a much smaller number of recent studies on the estimation of preferences that are nonseparable over time. In principle relaxing assumptions about preferences might seem an easier task than relaxing the CCM assumption, because at least variables affecting utility, such as leisure and consumption, are observed in panel data, whereas reliable data on the asset positions of people are typically much harder to obtain. Nevertheless allowing for aggregate shocks (transmitted through prices via markets) is quite challenging, and there is clearly room for further work in designing structural models with parameters that can be identified with computationally tractable estimators.

7.1 Applying Additivity in a Limited Way

Although CCM provides a useful framework for thinking about the data, like any null the failure to reject it may merely reject poor data, while a rejection should be interpreted as a strike against all the assumptions that bolster the framework, including those that are entirely unrelated to questions of market structure. Faced with this weak way of interpreting the evidence, a natural direction for future research is to relax CCM by explicitly modeling the rationale for incomplete markets, or relax the strong separability assumptions about preferences that underpin (6.87). However, tests of additivity can at least be adapted to particular markets, or specific populations. From (6.86) it is clear that only markets for which data exist are being tested; consequently this test has no power against commodities that are not traded.[18] Another way of estimating a more limited set of complete markets is to assume the gains from trade are only exhausted within a subpopulation, rather than over the whole. To see how this works we index by i the subpopulation in question. Then (6.86) holds if we subscript prices by i, and for any two households m and n in the same subpopulation i:

$$\Delta \ln [\partial u_t(c_{mt}, z_{mt})/\partial c_{1mt}] = \Delta \ln [\lambda_{1it}] = \Delta \ln [\partial u_t(c_{nt}, z_{nt})/\partial c_{1nt}]. \qquad (6.92)$$

Estimation proceeds in a similar manner as before, either through first differencing across individuals in the same subpopulation (Hayashi, Altonji, and Kotlikoff, 1996), or by estimating time dummies for each subpopulation (Miller and Sieg, 1997).

7.2 Incomplete Markets

A much weaker assumption than time additivity, yielding a more general framework, would be to relax this very specialized assumption about market structure, and assume only that agents sequentially allocate their resources to maximize their utility, for example by trading their labor for goods which are currently consumed for dividends or interest bearing asset claims. In this more general framework the individual or household objectives remain the same, but the opportunities the household faces are more complicated. With reference to the recursive formulation developed in section 6, rather than facing markets for contingent securities at prices $A_t(z^* \mid z)$ defined in (6.75), the agent can only buy or sell combinations of state contingencies in proportion to the payoffs of the available securities. Therefore instead of trading claims to wealth next period, only (a limited number of types of) securities are traded. For example if there are L (finite) states and a smaller number of securities, the agent expects to receive less utility over her lifetime than when the set of securities span the state space (except in certain cases like those mentioned in section 6 in which consumer preferences satisfy special aggregation conditions).

Let $q_{rnt} \equiv q_{rnt}(z)$ denote the quantity of asset r held by household n in period t when state z occurs, $p_{rt} \equiv p_{rt}(z)$ its price, and $s_{rt} \equiv s_{rt}(z)$ the associated dividend. In addition to constraining bequests (to be nonnegative, for example), the agent must satisfy, for each date t and state z, the inequalities:

$$\Sigma_{r \in R} p_{rt}(q_{rnt} - q_{rn,t-1}) + q_{rnt}s_{rt} \geq c_{nt}. \qquad (6.93)$$

This framework, which (as mentioned) is often referred to as the permanent

income hypothesis (PIH), yields two kinds of first-order conditions.[19] First is the spot market for goods consumed concurrently, examined in section 6. The other first-order condition for this problem equates the price of a good to be received next period in several states of the world, that is relative to its current spot price, with its matching marginal rate of substitution. Letting:

$$\pi_{rt} \equiv \pi_r(z_t) \equiv [p_r(z_{t+1}) + s_r(z_{t+1})]/p_r(z_t) \qquad (6.94)$$

denote the return on asset r in time t, the marginal rate of substitution between current consumption and future consumption is:

$$E\left\{\pi_{rt} \frac{\partial u_t(c_{nt+1}, z_{n,t+1})/\partial c_{1nt+1}}{\partial u_t(c_{nt}, z_{nt})/\partial c_{1nt}} \mid z_{nt}\right\} = 1, \qquad (6.95)$$

where, as before, $\partial u_t(c_{nt}, z_{nt})/\partial c_{1nt}$ is the marginal utility of the first good. Interpreting (6.95), note that if π_{rt} paid off two units in good states of the world, and nothing in bad states, (6.95) would be an equilibrium condition for the marginal rate of substitution between a unit of consumption for today versus several commodities which relate to consumption tomorrow, that bundle comprising two units which would be delivered in the good states alone. Unless an asset r exists (can be constructed) which pays off in only one state of the world, it is impossible for an optimally behaving consumer to form pairwise allocations, that depend only on her wealth and prices, between a good consumed in that particular state and the same good consumed now. Limited market opportunities compel her to buy bundles of commodities that are typically intertemporally linked through the fortunes of the assets she holds, even if her own utility function is additively separable across time.[20]

Several studies attempt to test the PIH with panel data based on equation (6.95). In the spirit of this literature, let ε_{1rnt} denote the difference between the expectation on the left side of (6.95) and its realization:

$$\varepsilon_{1rnt} \equiv E\left\{\pi_{rt} \frac{\partial u_t(c_{nt+1}, z_{n,t+1})/\partial c_{1nt+1}}{\partial u_t(c_{nt}, z_{nt})/\partial c_{1nt}} \mid z_{nt}\right\} - \pi_{rt} \frac{\partial u_t(c_{nt+1}, z_{n,t+1})/\partial c_{1nt+1}}{\partial u_t(c_{nt}, z_{nt})/\partial c_{1nt}}. \qquad (6.96)$$

Thus $E[\varepsilon_{1rnt} \mid z_{nt}] = 0$. Assuming the realization $\varepsilon_{1rnt} < 1$ for all (r, n, t), we obtain in logarithmic form the following identity for all (r, n, t) and z_{nt} by substituting (6.96) into (6.95):

$$\Delta \ln [\partial u_t(c_{nt}, z_{nt})/\partial c_{1nt}] = \ln (1 - \varepsilon_{1rnt}) - \ln (\pi_{rt}). \qquad (6.97)$$

Suppose we now average (6.97) over the cross-section (assumed to be a random sample) and define $\upsilon_{1rt}^{(0)}$ as the limit in the sample size N. That is:

$$\upsilon_{1rt}^{(0)} \equiv \operatorname*{plim}_{N\to\infty} N^{-1} \sum_{n=1}^{N} \ln (1 - \varepsilon_{1rnt}) - \ln (\pi_{rt})$$

$$= \operatorname*{plim}_{N\to\infty} N^{-1} \sum_{n=1}^{N} \Delta \ln [\partial u_t(c_{nt}, z_{nt})/\partial c_{1nt}]. \qquad (6.98)$$

The estimation approach is to treat $v_{1n}^{(0)}$ as a time dummy to be estimated in the second line of (6.98) and form an orthogonality condition in a cross-section or panel context. Similarly, one could multiply equation (6.97) by an instrument vector $y_{nt} = (y_{nt}^{(1)}, \ldots, y_{nt}^{(q)})'$, define a vector of dummy variables v_{1n} as:

$$v_{1rt} = \lim_{N\to\infty} N^{-1} \sum_{n=1}^{N} y_{nt}[\ln(1 - \varepsilon_{1tnt}) - \ln(\pi_{n})]. \qquad (6.99)$$

and obtain q orthogonality conditions of the form:

$$\lim_{N\to\infty} N^{-1} \sum_{n=1}^{N} y_{nt} \Delta \ln[\partial u_t(c_{nt}, z_{nt})/\partial c_{nt}] - v_{1n} = 0. \qquad (6.100)$$

where it is now assumed (as in the analogous time series literature) that all the arguments in $u_t(c_{nt}, z_{nt})$ are observed. Since the q-dimensional coefficient on the time dummy vector itself must be estimated if (6.100) is to be exploited in estimation, this in itself is of no help in identifying the structural parameters of interest, unless they are restricted in some way. For without such restrictions, there are more parameters to estimate than orthogonality conditions! As a practical matter, the studies essentially assume:

$$v_{1rt} = v_{1rt}^{(0)} \operatorname{plim}_{N\to\infty} N^{-1} \sum_{n=1}^{N} y_{nt}. \qquad (6.101)$$

Substitute (6.101) into (6.100) and minimize a quadratic criterion function in:

$$N^{-1} \sum_{n=1}^{N} y_{nt}[\Delta \ln[\partial u_t(c_{nt}, z_{nt})/\partial c_{1nt}] - v_{1rt}^{(0)}] \qquad (6.102)$$

by choosing one time dummy coefficient per year and the structural parameters of interest.[21] From (6.99), assumption (6.101) is equivalent to:

$$\operatorname{plim}_{N\to\infty} N^{-1} \sum_{n=1}^{N} y_{nt} \ln(1 - \varepsilon_{1tnt}) = \operatorname{plim}_{N\to\infty} N^{-1} \sum_{n=1}^{N} \ln(1 - \varepsilon_{1tnt}) \operatorname{plim}_{N\to\infty} N^{-1} \sum_{n=1}^{N} y_{nt}. \qquad (6.103)$$

There is no reason to believe that the time dummies obey the restrictions (6.101), or equivalently that instruments exist which satisfy (6.103) and are thus uncorrelated with the forecast error made by a cross-section of the population at a point in time. When markets are incomplete agents are forced to make choices over bundles or groups of goods that do not necessarily match the proportions they would desire if somehow the commodities were available separately. Thus one would expect past allocations, and also those factors which affected past choices, to affect current and future goods allocations as well. The very fact that the consumer cannot distinguish her choices over current quantities from what is available to her in the future even after controlling for her endowment now and her planned bequests is the reason why picking instruments is so difficult for econometricians. Thus it does not seem possible to identify and estimate parameters characterizing the utility function without imposing more assumptions than the PIH does.[22] Faced with this conundrum, recent empirical research on dynamic structural models has become more explicit about the ways in which markets might be incomplete. For example Rust and Phelan (1995) examine the effects of social security on retirement; Margiotta and Miller (1994) estimate a model of moral hazard and the design of optimal compensation plans for managers.

7.3 Nonseparable Preferences

This section concludes the analysis by considering some implications of relaxing the assumption that preferences over continuous choices are time separable. Notice that because (6.11) is solved for all K discrete choices, it must hold for the choice actually made. Therefore much of the discussion which follows applies more generally, even though discrete choices are not treated explicitly. To avoid the difficulties associated with identifying a model of incomplete markets we will also assume that the CCM budget constraint (6.76) applies.[23] Thus (6.73) pertains but (6.74), the assumption that the transition law does not depend on choices, is relaxed. Let η denote the Lagrange multiplier associated with the lifetime budget constraint (6.76). After dropping the k subscripts, the FOC (6.11) simplifies to:

$$\frac{\partial u_t(c_t, z)}{\partial c_{1t}} + \lambda_t(z)\eta$$
$$= -\int \frac{\partial v_{t+1}(z^*)}{\partial z^*} \frac{\partial g(s, c_t, z)}{\partial c_{1t}} h_t(s \mid c_t, z) ds - \int v_{t+1}(z^*) \frac{\partial h_t(s \mid c_t, z)}{\partial c_{1t}} ds. \qquad (6.104)$$

The difference between (6.104) and the analogous FOC in the time additive case studied in the previous section is the existence of the two expressions on the right side. While η, the unobserved marginal utility of wealth, could be eliminated by taking logarithms and first differencing to obtain the counterparts to (6.81) and (6.87), this generates a numerically messy expression inside a logarithm. An alternative is to partially difference across consecutive periods. Accordingly, define $\Delta(\lambda_{t+1}/\lambda_t)$, the partial difference operator, as:

$$\Delta(\lambda_{t+1}/\lambda_t) f(z_t) \equiv f(z_{t+1}) - (\lambda_{t+1}/\lambda_t) f(z_t) \qquad (6.105)$$

for any real vector mapping $f(z_t)$. Applying $\Delta(\lambda_{t+1}/\lambda_t)$ to (6.104) yields:

$$0 = \Delta\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left[\frac{\partial u_t(c_t, z)}{\partial c_{1t}} + \int \left\{\frac{\partial v_{t+1}(z^*)}{\partial z^*} \frac{\partial g(s, c_t, z)}{\partial c_{1t}} h_t(s \mid c_t, z) + v_{t+1}(z^*) \frac{\partial h_t(s \mid c_t, z)}{\partial c_{1t}}\right\} ds\right]. \qquad (6.106)$$

Apart from the nonlogarithmic form, the two expressions inside the integral of (6.106) distinguish it from (6.87), and both of them complicate estimation considerably. When c_{1t} affects the transition probability density $h_t(\varsigma \mid c_t, z)$ in a nondegenerate way, then an approximation for the valuation function $v_t(z)$, is required to evaluate the expression inside the integral. The first problem is, then, how to obtain such an approximation (and incorporate the approximation error into the asymptotic distribution theory for the estimation procedure). Note too that $v_t(z)$ and its derivative $\partial v_t(z)/\partial z$ must be evaluated at different points of the support for z. Also $h_t(\varsigma \mid c_t, z)$ and $g(\varsigma, c_t, z)$: in other words, $F_t(z \mid z_t, c_t)$ must be known or estimated. The second problem is how knowledge about $F_t(z \mid z_t, c_t)$ can be obtained from a cross-section or panel when aggregate fluctuations are not fully anticipated.

When all uncertainty about the future is idiosyncratic to individuals, the state variables are affected by the past choices in a deterministic way, and all time dependence in the state is observed, then equation (6.106) can easily be adapted for estimation purposes, as follows.[24] Suppose $z_{nt} \equiv (x_{nt}, \varepsilon_{nt})$, where as before x_{nt} are observed variables and ε_{nt} is unobserved. We further assume that the law of motion for x_{nt} is simply a mapping of (x_{nt}, c_{nt}), denoted $x_{n,t+1} \equiv \bar{g}(x_{nt}, c_{nt})$, not to be confused with the closely related but differently defined $g(\cdot)$ in (6.104) and (6.106), and that ε_{nt} is identically and independently distributed across (n, t) with probability distribution function $G(\varepsilon_{nt})$. Because aggregate shocks are fully anticipated by the population, the one-period interest rate r_t is $\lambda_{t-1}/\lambda_t - 1$, and:

$$0 = \Delta(1 + r_t)\left[\frac{\partial u_t(c_{nt}, x_{nt}, \varepsilon_{nt})}{\partial c_{1nt}} + \int \frac{\partial v_{t+1}(x_{n,t+1}, \varepsilon)}{\partial x_{n,t+1}} \frac{\partial \bar{g}(x_{nt}, c_{nt})}{\partial c_{1nt}} dG(\varepsilon)\right]. \tag{6.107}$$

Expressions for derivatives of the valuation function are found by applying the envelope theorem to $v_t(x_{nt}, \varepsilon_{nt})$, and the chain rule for differentiating composite functions, to obtain:

$$\int \left[\frac{\partial v_{t+1}(x_{n,t+1}, \varepsilon)}{\partial x_{n,t+1}} \frac{\partial \bar{g}(x_{nt}, c_{nt})}{\partial c_{1nt}}\right] dG(\varepsilon)$$
$$= \sum_{s=t+1}^{T} E\left[\frac{\partial u_s(c_{ns}, x_{ns}, \varepsilon)}{\partial x_{ns}} \frac{\partial \bar{g}(x_{nt}, c_{nt})}{\partial c_{1nt}} \prod_{r=t+1}^{s-1} \frac{\partial \bar{g}(x_{nr}, c_{nr})}{\partial x_{nr}} \mid x_{nt}, c_{nt}\right], \tag{6.108}$$

where the expectation on the right side of (6.108) is taken over future idiosyncratic shocks and their effects on future consumption choices. The population moments and their corresponding sample analogs for use in estimation are based on equations (6.107) and (6.108). Up to some $\theta_0 \in \Theta$ to be estimated, suppose $\bar{g}(x_{nt}, c_{nt})$ can be paremeterized as a known function $\bar{g}(x_{nt}, c_{nt}, \theta_0)$, and the marginal utility of consumption is also linear in the unobserved variables. That is to say:

$$u_t(c_{nt}, x_{nt}, \varepsilon_{nt}) = \kappa_t(c_{nt}, x_{nt}, \theta_0) + c_{nt}\varepsilon_{nt} \tag{6.109}$$

for some mapping $\kappa_t(c_{nt}, x_{nt}, \theta_0)$ known up to $\theta_0 \in \Theta$. Assuming there exists an instrument vector y_{nt} satisfying the condition $E[y_{nt}\varepsilon_{ns}] = 0$ for all s greater than t, let $f_n(\theta) \equiv y_n\Delta(1 + r_t)h_n(\theta)$, where:

$$h_n(\theta) = \frac{\partial \kappa_t(c_{nt}, x_{nt}, \theta)}{\partial c_{1nt}}$$
$$+ \sum_{s=t-1}^{r} \frac{\partial \kappa_s(c_{ns}, x_{ns}, \theta)}{\partial x_{ns}} \frac{\partial \bar{g}(x_{nt}, c_{nt}, \theta)}{\partial c_{1nt}} \prod_{r=t+1}^{s-1} \frac{\partial \bar{g}(x_{nr}, c_{nr}, \theta)}{\partial x_{nr}}. \tag{6.110}$$

Then equations (6.107) through (6.110) imply $E[f_n(\theta)]$ has a root at θ_0, which is now assumed to be unique in Θ. Following the procedure outlined in equations (6.85) and (6.86) then yields a $N^{1/2}$ consistent and asymptotically normal estimator for θ_0.

To illustrate, suppose there is only one good, the vector c_t collapsing to a positive real number c_{1t}, and current consumption is additively separable from everything apart from the previous period's consumption. More specifically, assume:

$$\kappa_t(c_t, x_t, \theta) = \beta^t(c_t^{\alpha} + \gamma c_t c_{t-1}). \tag{6.111}$$

If there are no other observed state variables, then $x_t = c_{t-1}$ and $\bar{g}(\cdot)$ is the identity function. Ignoring the factor of proportionality β^t which plays no role in estimation:

$$h_n(\theta) = \alpha c_{nt}^{\alpha-1} + \gamma c_{n,t+1}$$
$$f_n(\theta) = y_n[\beta(\alpha c_{n,t+1}^{\alpha-1} + \gamma c_{n,t+2}) - (1 + r_t)(\alpha c_{nt}^{\alpha-1} + \gamma c_{n,t+1})]. \tag{6.112}$$

Noting $N^{-1}\Sigma_{n=1}^{N} f_n(\theta_0)$ converges to 0 in probability thus provides a basis for the estimation of $\theta_0 = (\alpha, \gamma)$ with a household panel sampled over two periods.

8 Conclusion

This chapter has studied how econometricians confront cross-sectional and panel data to make inferences about the dynamic optimization problems that rational individuals solve, and the problems that an econometrician faces when all the respondents in the sample participate in the same competitive equilibrium. Clearly considerable progress has been made in establishing the conceptual links between theoretical abstractions and their empirical counterparts. Two benefits from this greater sophistication are that the estimated models are easy to interpret, and that models not explicitly based on economic theory are now more susceptible to criticism than before. However, other benefits touted for structural modeling have failed to materialize, or have yielded an unintended byproduct. When theories are only loosely connected to data analysis, overfitting is possible, creating room for expansive claims. Structural modeling imposed a discipline on empirical research that led to a rude awakening: theories which appear plausible and seem to agree with broad empirical regularities receive considerably less support when subjected to the rigors of structural modeling. Scaling back bold claims about the data to accommodate the findings of dynamic structural models is a healthy tonic for the profession, encouraging us to search for bigger, more informative data sets and to develop more subtle hypotheses.

Nevertheless, inferring structural parameters of dynamic models from micro data has not yet gained widespread acceptance within the profession. Many years after the conception of this exciting idea, some economists still question whether the value of a seamless transition between theory and econometric practice is worth the effort. First, the reductionism, that economists can fruitfully empirically analyze behavior as intimate as fertility and marriage by applying mathematics and statistics to sample records of real people, is anathema to many who might otherwise feel comfortable with the broad postulates of economics thinking, such as rational behaviour, and are even willing to muse over theories of household production. Second are the rigid parameterizations to which applied econometricians limit themselves, to achieve the twin goals of internal consistency and tractability, at the expense of discarding more flexible functional forms; similarly imposing the optimization and equilibrium postulates is met with skepticism despite the absence of convincing alternatives. Third, although micro data sets are much more detailed and contain greater numbers of observations than ever before, they still lack relevant sample information econometricians would incorporate if they could. Fourth, unobserved heterogeneity to accommodate deficiencies in the data is typically modeled in a rudimentary way, with the aim of achieving a parsimonious, empirically tractable specification, rather than addressing any specific data limitation. Fifth, much work on dynamic discrete choice is estimated using ML so that (except for further nesting within the overall framework via further restrictions on the parameter space) there is no sample information left over to test the specification; in addition those models that are estimated with GMM yield overidentifying restrictions that do not usually point to an interesting alternative economics framework. The sixth limitation is attributable to the high cost of programming nonlinear models: revisiting data sets to verify previous results and check their sensitivity to minor specification changes is hard work. The net result of these drawbacks is that one desirable feature sometimes claimed for structural estimation, that it delivers policy invariant parameters which can be used in formulating policy advice, is not yet taken very seriously by policy makers themselves. For this reason alone, these limitations present formidable challenges to future research, which practitioners building dynamic structural models for estimation purposes must meet if the area is to prove more durable than a fad.

Paradoxically, the goal of meeting this policy relevance criterion may require the tasks of estimation and inference to be split from policy evaluation. Some of the earliest applications in dynamic structural modeling joined policy analysis and estimation together, by solving the value function for lots of vectors in the paremeter space to implement ML, and then conducting policy analysis by perturbing the estimated parameters. But in the last decade, faster computers and tremendous advances in telecommuncation technologies have not only brought richer data sets with greater numbers of observations, and speedier processors. Researchers have become more keenly aware of the vast gulf separating our propensity as economic theorists to write down dynamic structural models, and our ability as practitioners in numerical analysis to solve them. By only solving structural models for parameters that have first been estimated from identifying equations that apply cheaply computed IV estimators, perhaps we can more easily capitalize on the huge microeconomics data tracts that are becoming increasingly available in estimation, and concentrate our numerical analysis on models that make economic sense to policy makers.

Notes

Without implicating anybody, I thank Xiaorong Dong, Yaniv Grinstein, and Nitin Mehta for research assistance, and Mico Mrkaic, Vesna Prasnikar, John Rust, and Holger Sieg for comments on an earlier draft.

1 This field is no exception to the rule that surveys and commentary accompany its growth. Reviews of the estimation of dynamic models of discrete choice have been undertaken by Eckstein and Wolpin (1989a), Pakes (1994), and Rust (1994a; 1994b). The monograph by Devine and Kiefer (1991) and the chapter by Neumann in this volume provide excellent coverage of the search literature, to which only passing reference will be made here. In addition to the chapters in this volume on labor supply, consumer demand systems, and production, readers are referred to Altug and Labadie (1994), Browning and Luscardi (1996), and Miller and Sieg (1997) on the equilibrium aspects of continuous choices. A nontechnical introduction to the critical role panel data have played in advancing this literature is given in Miller (1993).

2 Readers interested in a more technical and more extensive survey of this particular topic are referred to Rust (1995). See also the discussion and references contained in chapter 7 in volume I of this handbook by Kim and Pagan (1995).

3 The literature on the rational expectations hypothesis is vast. For example see Muth (1961) for an early statement and rationale, Lucas (1976) on its implications for policy advisers, and Pesaran (1987) for a detailed critique.

4 For example Pakes (1994) notes "that not all models ... generate Euler equations from simple compensating variations in adjacent periods ... Here we ... [illustrate] the logic of the argument that allows one to use Euler equations when they are available" (p. 186).

5 Several studies relax the distributional assumptions about unobserved heterogeneity, but in practice assume a probability distribution with mass at a finite number of

points for the purposes of statistical inference and hypothesis testing (Miller, 1984; Tabor, 1995).

6 If Z is a vector space rather than a finite set, and $F_k(z_n^* \mid z_n, \theta)$ is the cumulative probability distribution function which supports a probability density function that is differentiable in each $\theta \in \Theta$, then a similar log-likelihood can of course be formed using the density, instead of the probability, in equation (6.12).

7 This more ambitious research agenda attempts to free unobserved heterogeneity from assumptions about parametric functional forms, but the price of this requirement is to greatly reduce the class of models that are identified. See Elbers and Ridder (1982) and Heckman and Singer (1984) for analyses of duration models which are semiparametrically identified.

8 For example the model should not be rejected by the information matrix test (White, 1982).

9 See Zellner (1971), for example, on the Bayesian updating used in the derivation of (6.17) and (6.18).

10 Thus the subscript on $v_1(\cdot)$ indicates $d_t = 1$ and is not a time subscript, which would be redundant in a time stationary Markov problem.

11 Miller (1984) used a cubic spline to interpolate between the points at which the integrand is computed, although other methods are available. See Powell (1981), for example, for an introduction to approximation theory.

12 See also Keane (1994) and Geweke and Keane (1996), and the articles referenced there, for more detailed analysis and Monte Carlo evidence on using simulation techniques to estimate models with limited dependent variables and serially correlated unobservables with panel data. The latter paper takes a Bayesian approach to estimating dynamic programming models.

13 See Smart (1974) for a discussion of the contraction mapping theorem, and the related approximation theorem. Stokey and Lucas (1989) and Altug and Labadie (1994) also contain extensive discussions of applications of the contraction mapping in economics.

14 The probability distribution function for ε_{1t} with location parameter ϑ is $\exp(-e^{(\varepsilon_{1t}+\vartheta)})$.

15 See McFadden (1973) or Maddala (1984), for example.

16 Hotz and Miller (1993) and Altug and Miller (1997) also consider models where observed variables in the state space belong to a vector space. The choice and outcome probabilities are estimated nonparametrically in this case, but the resulting semiparametric estimator for the structural parameters retain $N^{1/2}$ consistency and asymptotic normality.

17 See Mace (1991), Cochrane (1991), Townsend (1994), Udry (1994), Ham and Jacobs (1994), as well as Hayashi, Altonji, and Kotlikoff (1996).

18 For example, the findings of Altug and Miller (1990) should be interpreted as statements about contingent allocations of labor and food, rather than about all markets.

19 Hall and Mishkin (1982) began this literature. Recent contributions include Altonji and Siow (1987), Zeldes (1989), Runkle (1991), Mariger and Shaw (1993), Nelson (1994), and Luscardi (1996).

20 Altug and Miller (1990) elaborate on this point and its econometric implications using the lifetime budget formulation (pp. 548–50). In chapter 9 of their monograph Altug and Labadie (1994) illustrate it with a borrowing constraint. This subsection draws heavily from Miller and Sieg (1997). See also Card (1994) for a discussion of this identification problem.

21 For example Hayashi, Altonji, and Kotlikoff (1996) recently asserted that: "Although, as first pointed out by Chamberlain (1984) and subsequently emphasized by a number of authors, a zero time-series correlation does not necessarily imply a zero cross-section correlation, the cross-section correlation will also be zero if the stochastic environment can be represented as the sum of a macro component common to all households and an idiosyncratic component" (p. 271), but they provide neither references nor conditions for when this additive condition holds.

22 Recognizing the difficulties associated with finding such instruments, Zeldes (1989) and Runkle (1991) have taken a second-order expansion of the error term, but unfortunately this approach does not overcome the problems discussed here.

23 Recently several studies have integrated continuous and discrete choices (Aguirregabria, 1994; Altug and Miller, 1997).

24 See Shaw (1989) for an application of this approach to human capital accumulation through experience on the job.

References

Aguirregabria, V. (1994), "Dynamic Models of Price and Inventory Decisions with Fixed Adjustment Costs," PhD dissertation, Universidad Complutense de Madrid.

Ahn, N. (1995), "Measuring the Value of Children by Sex and Age Using a Dynamic Programming Model," *Review of Economic Studies*, 62, 361–379.

Altonji, A. and A. Siow (1987), "Testing the Response of Consumption to Income Changes with (Noisy) Panel Data," *Quarterly Journal of Economics*, 102, 293–328.

Altug, S. and P. Labadie (1994), *Dynamic Choice and Asset Markets*, Academic Press.

Altug, S. and R. Miller (1990), "Household Choices in Equilibrium," *Econometrica*, 58, 542–570.

Altug, S. and R. Miller (1997), "The Effect of Work Experience on Female Wages and Labor Supply," *Review of Economic Studies*, forthcoming.

Arrow, K. (1963), "The Role of Securities in the Optimal Allocation of Risk-Bearing," *Review of Economic Studies*, 45, 96–96.

Bellman, R. (1957), *Dynamic Programming*, Princeton University Press.

Berkovec, J. and S. Stern (1991), "Job Exit Behavior of Older Men," *Econometrica*, 59, 189–210.

Boswell, D. (1994), "A Dynamic Stochastic Model of Medical Care Use and Work Absence." University of Minnesota Working Paper.

Browning, M. and A. Luscardi (1996), "Household Saving: Micro Theories and Micro Facts," *Journal of Economic Literature*, 34, 1797–1855.

Card, D. (1994), "Intertemporal Labour Supply: An Assessment," in *Advances in Econometrics: Sixth World Congress*, ed. C. Sims, Cambridge University Press, chapter 2, 49–80.

Chamberlain, G. (1984), "Panel Data," in *Handbook in Econometrics*, vol. 2, eds Z. Griliches and M. Intriligator, North Holland, chapter 22.

Cochrane, J. (1991), "A Simple Test of Consumption Insurance," *Journal of Political Economy*, 99, 957–977.

Das, M. (1992), "A Micro-Econometric Model of Capital Utilization and Retirement: The Case of the U.S. Cement Industry," *Review of Economic Studies*, 59, 277–298.

Debreu, G. (1959), *Theory of Value*, Yale University Press.

Devine, T. and N. Kiefer (1991), *Empirical Labor Economics: The Search Approach*, Oxford University Press.

Eckstein, Z. and K. Wolpin (1989a), "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey," *Journal of Human Resources*, 24, 562–598.

Eckstein, Z. and K. Wolpin (1989b), "Dynamic Labour Force Participation of Married Women and Endogenous Workforce Experience," *Review of Economic Studies*, 56, 375–390.

Elbers, C. and G. Ridder (1982), "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, 49, 403–410.

Flinn, C. (1986), "Wages and Job Mobility of Young Workers," *Journal of Political Economy*, 94, S88–S110.

Flinn, C. and J. Heckman (1982), "New Methods for Analysing Structural Models of Labor Force Dynamics," *Journal of Econometrics*, 18, 115–168.

Flyer, F. (1995), "The Influence of Higher Moments of Earning Distributions on Career Decisions," *Journal of Labor Economics*, forthcoming.

Gallant, A. and D. Jorgenson (1979), "Statistical Inference for a System of Simultaneous, Nonlinear, Implicit Equations in the Context of Instrumental Variables Estimation," *Journal of Econometrics*, 11, 272–302.

Geweke, J. and M. Keane (1996), "Bayesian Inference for Dynamic Discrete Choice Models without the Need for Dynamic Programming," Federal Reserve Bank of Minneapolis Research Department Working Paper.

Hall, R. and F. Mishkin (1982), "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households," *Econometrica*, 50, 461–481.

Ham, J. and K. Jacobs (1994), "Using Exogenous Information on Sector Specific and Idiosyncratic Shocks to Test for Full Insurance," University of Pittsburgh Working Paper.

Hansen, L. and K. Singleton (1982), "Generalized Instrumental Variables Estimators of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

Hayashi, F., J. Altonji and L. Kotlikoff (1996), "Risk Sharing between and within Families," *Econometrica*, 64, 261–294.

Hansen, L. (1982), "Large Sample Properties of Generalized Methods of Moments Estimators," *Econometrica*, 50, 1029–1054.

Heckman, J. (1981a), "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Data with Econometric Applications*, eds C. Manski and D. McFadden, MIT Press, chapter 3, 114–178.

Heckman, J. (1981b), "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time–Discrete Data Stochastic Process," in *Structural Analysis of Discrete Data with Econometric Applications*, eds C. Manski and D. McFadden, MIT Press, chapter 4, 179–195.

Heckman, J. and T. MaCurdy (1980), "A Life Cycle Model of Female Labor Supply," *Review of Economic Studies*, 47, 47–74.

Heckman, J. and B. Singer (1984), "A Method of Minimizing the Impact of Distributional Assumptions in Econometric Models of Duration Data," *Econometrica*, 52, 271–320.

Hotz, V., F. Kydland and G. Sedlacek (1988), "Intertemporal Preferences and Labor Supply," *Econometrica*, 56, 335–360.

Hotz, V. and R. Miller (1988), "An Empirical Analysis of Life Cycle Fertility and Female Labor Supply," *Econometrica*, 56, 91–118.

Hotz, V. and R. Miller (1993), "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 60, 497–529.

Hotz, V., R. Miller, S. Sanders and G. Smith (1994), "A Simulation Estimator for Dynamic Models of Discrete Choice," *Review of Economic Studies*, 61, 265–289.

Jovanovic, B. (1979), "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 87, 972–990.

Keane, M. (1994), "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, 62, 95–116.

Keane, M. and K. Wolpin (1994a), "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence," *Review of Economics and Statistics*, 76, 648–672.

Keane, M. and K. Wolpin (1994b), "The Career Decisions of Young Men," University of Minnesota and New York University Working Paper.

Kiefer, N. and G. Neumann (1979), "An Empirical Job-Search Model with a Test of the Constant Reservation Wage Hypothesis," *Journal of Political Economy*, 87, 89–108.

Kiefer, N. and G. Neumann (1981), "Individual Effects in a Nonlinear Model: Explicit Treatment of Heterogeneity in the Empirical Job Search Model," *Econometrica*, 49, 965–979.

Kim, K. and A. Pagan (1995), "The Econometric Analysis of Calibrated Macroeconomic Models," in *Handbook of Applied Econometrics. Volume I: Macroeconomics*, eds M. Hasham Pesaran and M. Wickens, Blackwell, 356–390.

Lucas, R. (1976), "Econometric Policy Evaluation: A Critique," *Carnegie-Rochester Conference Series*, 1, 19–46.

Luscardi, A. (1996), "Permanent Income, Current Income and Consumption: Evidence from Two Panel Data Sets," *Journal of Business and Economic Statistics*, 14, 81–90.

Mace, B. (1991), "Full Insurance in the Presence of Aggregate Uncertainty," *Journal of Political Economy*, 99, 928–956.

MaCurdy, T. (1981), "An Empirical Model of Labor Supply in a Life-Cycle Setting," *Journal of Political Economy*, 89, 1059–1085.

MaCurdy, T. (1983), "A Simple Scheme for Estimating an Intertemporal Model of Labor Supply and Consumption in the Presence of Taxes and Uncertainty," *International Economic Review*, 24, 265–289.

Maddala, G. (1984), *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press.

Margiotta, M. and R. Miller (1994), "Managerial Compensation and the Cost of Moral Hazard," Carnegie Mellon University Working Paper.

Mariger, R. and K. Shaw (1993), "Unanticipated Aggregate Disturbances and Tests of the Life Cycle Model Using Panel Data," *Review of Economics and Statistics*, 75, 48–56.

McFadden, D. (1973), "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, Academic Press.

McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Estimation," *Econometrica*, 57, 995–1027.

Miller, R. (1982), "Job Specific Capital and Labor Mobility," PhD dissertation, University of Chicago.

Miller, R. (1984), "Job Matching and Occupational Choice," *Journal of Political Economy*, 92, 1086–1120.

Miller, R. (1993), "What Can Panel Data Tell Us about Labour Markets," *Labour Economics and Productivity*, 5, 165–180.

Miller, R. and P. Ramnath (1995), "Plant Objectives as Revealed by Shop Floor Activities," *Journal of Operations Management*, 12, 273–296.

Miller, R. and H. Sieg (1997), "A Microeconometric Comparison of Household Behavior between Countries," *Journal of Business and Economic Statistics*, 15, 237–254.

Miranda, M. and G. Schnitkey (1995), "An Empirical Model of Asset Replacement in Dairy Production," *Journal of Applied Econometrics*, 10, S41–S45.

Miranda, M. and X. Rui (1996), "Maximum Likelihood Estimation of the Nonlinear Rational Expectations Asset Pricing Model," *Journal of Economic Dynamics and Control*, forthcoming.

Muth, J. (1961), "Rational Expectations and the Theory of Price Movements," *Econometrica*, 29, 315–335.

Nelson, J. (1994), "On Testing for Full Insurance Using Consumer Expenditure Data," *Journal of Political Economy*, 102, 384–394.

Miller, R. and S. Sanders (1996), "Human Capital Development and Welfare Participation," *Carnegie–Rochester on Public Policy*, 46, forthcoming.

Newey, W. (1984), "A Methods of Moments Interpretation of Sequential Estimators," *Economic Letters*, 15, 201–206.

Pakes, A. (1986), "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 54, 755–784.

Pakes, A. (1994); "Dynamic Structural Models, Problems and Prospects: Mixed Continuous Discrete Controls and Market Interactions," in *Advances in Econometrics: Sixth World Congress*, vol. II, ed. C. Sims, Cambridge University Press, chapter 5, 171–259.

Pakes, A. and D. Pollard (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1058.

Pesaran, M. (1987), *The Limits to Rational Expectations*, Blackwell.

Powell, M. (1981), *Approximation Theory and Methods*, Cambridge University Press.

Rosenzweig, M. and K. Wolpin (1993), "Credit Market Constraints, Consumption Smoothing, and the Accumulation of Durable Production Assets in Low-Income Countries: Investments in Bullocks in India," *Journal of Political Economy*, 101, 223–244.

Rothwell, G. and J. Rust (1995), "A Dynamic Programming Model of U.S. Nuclear Power Plant Operations," Stanford University Working Paper.

Rubenstein, M. (1981), "A Discrete-Time Synthesis of Financial Theory," *Research in Finance*, 3, 53–102.

Runkle, D. (1991), "Liquidity Constraints and the Permanent-Income Hypothesis: Evidence from Panel Data," *Journal of Monetary Economics*, 27, 73–98.

Rust, J. (1987), "Optimal Replacement of GMC Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 57, 999–1035.

Rust, J. (1994a), "Estimation of Dynamic Structural Models, Problems and Prospects: Discrete Decision Processes," in *Advances in Econometrics: 6th World Congress*, vol. II, ed. C. Sims, Cambridge University Press, chapter 4, 119–171.

Rust, J. (1994b), "Structural Estimation of Markov Decision Processes," in *Handbook of Econometrics*, vol. 4, eds R. Engle and D. McFadden, Elsevier Science, chapter 51, 3081–3143.

Rust, J. (1995), "Numerical Dynamic Programming in Economics," in *Handbook of Computational Economics*, eds H. Amman, D. Kendrick and J. Rust.

Rust, J. and C. Phelan (1995), "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," *Econometrica*, forthcoming.

Slade, M. (1994), "Optimal Pricing with Costly Adjustment and Persistent Effects: Empirical Evidence," University of British Columbia Working Paper.

Sieg, H. (1995), "Income Taxation, Household Choices and the Tax Burden: An Empirical Analysis," Carnegie Mellon University Working Paper.

Shaw, K. (1989), "Life-Cycle Labor Supply with Human Capital Accumulation," *International Economic Review*, 30, 431–456.

Smart, J. (1974), *Fixed Point Theorems*, Cambridge University Press.

Stock, J. and D. Wise (1990), "Pensions, the Option Value of Work and Retirement," *Econometrica*, 58, 1151–1180.

Stokey, N. and R. Lucas, with E. Prescott (1989), *Recursive Methods in Economic Dynamics*, Harvard University Press.

Tabor, T. (1995), "The Rising College Premium in the Eighties: Return to College or Return to Ability," University of Chicago Working Paper.

Trick, M. and S. Zin (1993), "A Linear Programming Approach to Solving Stochastic Dynamic Programs," Carnegie Mellon University Working Paper.

Townsend, R. (1994), "Risk and Insurance in Village India," *Econometrica*, 62, 539–591.

Udry, C. (1994), "Risk and Insurance in Northern Nigeria," *Review of Economic Studies*, 63, 495–526.

Van der Klaauw, W. (1993), "Female Labor Supply and Marital Status Decisions: A Life Cycle Model," New York University Working Paper.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–26.

Wolpin, K. (1984), "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality," *Journal of Political Economy*, 92, 1151–1180.

Wolpin, K. (1987), "Estimating a Structural Search Model: The Transition from School to Work," *Econometrica*, 55, 801–818.

Wolpin, K. (1992), "The Determinants of Black-White Differences in Early Employment Careers: Search, Layoffs, Quits, and Endogenous Wage Growth," *Journal of Political Economy*, 100, 535–560.

Zeldes, S. (1989), "Consumption and Liquidity Constraints: An Empirical Investigation," *Journal of Political Economy*, 97, 305–346.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, Wiley.