The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications

Michael P. Keane[*]   Petra E. Todd[†]   Kenneth I. Wolpin[‡]

March, 2010

ABSTRACT

The purpose of this chapter is twofold: (1) to provide an accessible introduction to the methods of structural estimation of discrete choice dynamic programming models (DCDP) and (2) to survey the contributions of applications of these methods to substantive and policy issues in labor economics. The first part of the chapter describes solution and estimation methods for DCDP models using, for expository purposes, a prototypical female labor force participation model. The next part reviews the contribution of the DCDP approach to three leading areas in labor economics: labor supply, job search and human capital. The final section discusses approaches to validating DCDP models.

---

[*]University of Technology, Sydney and Arizona State University

[†]University of Pennsylvania

[‡]University of Pennsylvania.

I. Introduction

The purpose of this chapter is twofold: (1) to provide an accessible introduction to the methods of structural estimation of discrete choice dynamic programming models (DCDP) and (2) to survey the contributions of applications of these methods to substantive and policy issues in labor economics.[1] The development of estimation methods for DCDP models over the last 25 years opened up new frontiers for empirical research in labor economics as well as other areas such as industrial organization, economic demography, health economics, development economics and political economy.[2] Reflecting the generality of the methodology, the first DCDP papers, associated with independent contributions by Gotz and McCall (1984), Miller (1984), Pakes (1986), Rust (1987) and Wolpin (1984), addressed a variety of topics, foreshadowing the diverse applications to come in labor economics and other fields. Gotz and McCall considered the sequential decision to re-enlist in the military, Miller the decision to change occupations, Pakes the decision to renew a patent, Rust the decision to replace a bus engine and Wolpin the decision to have a child.

The first part of this chapter provides an introduction to the solution and estimation methods for DCDP models. We begin by placing the method within the general latent variable framework of discrete choice analysis. This general framework nests static and dynamic models and nonstructural and structural estimation approaches. Our discussion of DCDP models starts by considering an agent making a binary choice. For concreteness, and for simplicity, we take as a working example the unitary model of a married couple's decision about the woman's labor force participation. To fix ideas, we use the static model with partial wage observability, that is, when wage offers are observed only for women who are employed, to draw the connection between theory, data and estimation approach. In that context, we delineate several goals of estimation, for example, testing theory or evaluating counterfactuals, and discuss the ability of alternative estimation approaches, encompassing those that are parametric or non-parametric and structural or non-structural, to achieve

---

[1] More technical discussions can be found in the surveys by Rust (1993, 1994), Miller (1997) and Aguirregebaria and Mira (forthcoming), as well as in a number of papers cited throughout this chapter.

[2] Their use has spread to areas outside of traditional economics, such as marketing, in which it is arguably now the predominant approach to empirical research.

those goals. We show how identification issues relate to what one can learn from estimation.

The discussion of the static model sets the stage for dynamics, which we introduce again, for expository purposes, within the labor force participation example by incorporating a wage return to work experience (learning by doing).[3] A comparison of the empirical structure of the static and dynamic models reveals that the dynamic model is, in an important sense, a static model in disguise. In particular, the essential element in the estimation of both the static and dynamic model is the calculation of a latent variable representing the difference in payoffs associated with the two alternatives (in the binary case) that may be chosen. In the static model, the latent variable is the difference in alternative-specific utilities. In the case of the dynamic model, the latent variable is the difference in alternative-specific value functions (expected discounted values of payoffs). The only essential difference between the static and dynamic cases is that alternative-specific utilities are more easily calculated than alternative-specific value functions, which require solving a dynamic programming problem. In both cases, computational considerations play a role in the choice of functional forms and distributional assumptions.

There are a number of modeling choices in all discrete choice analyses, although some are more important in the dynamic context because of computational issues. Modeling choices include the number of alternatives, the size of the state space, the error structure and distributional assumptions and the functional forms for the structural relationships. In addition, in the dynamic case, one must make an assumption about how expectations are formed.[4] To illustrate the DCDP methodology, the labor force participation model assumes additive, normally distributed, iid over time errors for preferences and wage offers. We first discuss the role of exclusion restrictions in identification and work through the solution and estimation procedure. We then show how a computational simplification can be achieved

---

[3]Most appliactions of DCDP models assume that agents, usually individuals or households, solve a finite horizon problem in discrete time. For the most part, we concentrate on that case and defer discussion of infinite horizon models to the discussion of the special case of job search models We do not discuss continuous time models except in passing.

[4]The conventional approach assumes that agents have rational expectations. An alternative approach directly elicits subjective expectations (see, e.g., Dominitz & Manski 1996, 1997; Van der Klaauw 2000; Manski 2004).

by assuming errors to be independent type 1 extreme value (Rust (1987)) and describe the model assumptions that are consistent with adopting that simplification. Although temporal independence of the unobservables is often assumed, the DCDP methodology does not require it. We show how the solution and estimation of DCDP models is modified to allow for permanent unobserved heterogeneity and for serially correlated errors. In the illustrative model, the state space was chosen to be of a small finite dimension. We then describe the practical problem that arises in implementing the DCDP methodology as the state space expands, the well-known curse of dimensionality (Bellman (1957)), and describe suggested practical solutions found in the literature including discretization, approximation and randomization.

To illustrate the DCDP framework in a multinomial choice setting, we extend the labor force participation model to allow for a fertility decision at each period and for several levels of work intensity. In that context, we also consider the implications of introducing non-additive errors (that arise naturally within the structure of models that fully specify payoffs and constraints) and general functional forms. It is a truism that any dynamic optimization model that can be (numerically) solved can be estimated.

Throughout the presentation, the estimation approach is assumed to be maximum likelihood or, as is often the case when there are many alternatives, simulated maximum likelihood. However, with simulated data from the solution to the dynamic programming problem, other methods, such as minimum distance estimation, are also available. We do not discuss those methods because, except for solving the dynamic programming model, their application is standard. Among the more recent developments in the DCDP literature is a Bayesian approach to the solution and estimation of DCDP models. Although the method has the potential to reduce the computational burden associated with DCDP models, it has not yet found wide application. We briefly outline the approach. All of these estimation methods require that the dynamic programming problem be fully solved (numerically). We complete the methodology section with a brief discussion of a method that does not require solving the full dynamic programming problem (Hotz-Miller (1993)).

Applications of the DCDP approach within labor economics have spanned most major areas of research. We discuss the contributions of DCDP applications in three main areas: (i)

labor supply, (ii) job search and (iii) schooling and career choices. Although the boundaries among these areas are not always clear and these areas do not exhaust all of the applications of the method in labor economics, they form a reasonably coherent taxonomy within which to demonstrate key empirical contributions of the approach.[5] In each area, we show how the DCDP applications build on the theoretical insights and empirical findings in the prior literature. We highlight the findings of the DCDP literatures, particularly those that involve counterfactual scenarios or policy experiments.

The ambitiousness of the research agenda that the DCDP approach can accommodate is a major strength. This strength is purchased at a cost. To be able to perform counterfactual analyses, DCDP models must rely on extra-theoretic modeling choices, including functional form and distributional assumptions. Although the DCDP approach falls short of an assumption-free ideal, as do all other empirical approaches, it is useful to ask whether there exists convincing evidence about the credibility of these exercises. In reviewing the DCDP applications, we pay careful attention to the model validation exercises that were performed. The final section of the chapter addresses the overall issue of model credibility.

II.   The Latent Variable Framework for Discrete Choice Problems

The development of the DCDP empirical framework was a straightforward and natural extension of the static discrete choice framework. The common structure they share is based on the latent variable specification, the building block for all economic models of discrete choice. To illustrate the general features of the latent variable specification, consider a binary choice model in which an economic agent with imperfect foresight, denoted by i, makes a choice at each discrete period t, from $t = 1, ..T$, between two alternatives $d_{it} \in \{0, 1\}$. In the labor economics context, examples might be the choice of whether to accept a job offer or remain unemployed or whether to attend college or enter the labor force. The outcome is determined by whether a latent variable, v_{it}^*, reflecting the difference in the (expected) pay-

---

[5] A notable omission is the literature on retirement behavior. Although that literature relies heavily on the DCDP approach, the previous Handbook of Labor Economics chapter by Lumsdaine and Mitchell provides an extensive survey.up to that time. We decided to concentrate on DCDP literatures that to date have not been surveyed in the Handbook.

offs of the $d_{it} = 1$ and $d_{it} = 0$ alternatives, crosses a scalar threshold value, which, without loss of generality, is taken to be zero. The preferred alternative is the one with the largest payoff, i.e., where $d_{it} = 1$ if $v_{it}^* \geq 0$ and $d_{it} = 0$ otherwise.

In its most general form, the latent variable may be a function of three types of variables: $\widetilde{D}_{it}$, a vector of the history of past choices $(d_{i\tau} : \tau = 1, ..., t-1)$, $\widetilde{X}_{it}$, a vector of contemporaneous and lagged values of J additional variables $(X_{ij\tau} : j = 1, ..., J; \tau = 1, ..., t)$ that enter the decision problem, and $\widetilde{\epsilon}_{it}$ $(\epsilon_{i\tau}: \tau = 1, ...t)$, a vector of contemporaneous and lagged unobservables that also enter the decision problem.[6] The agent's decision rule at each age is given by whether the latent variable crosses the threshold, that is,

$$
\begin{aligned}
d_{it} &= 1 \; if \;\; v_{it}^*(\; \widetilde{D}_{it}, \widetilde{X}_{it}, \widetilde{\epsilon}_{it}) \geq 0, \\
&= 0 \; if \;\; v_{it}^*(\; \widetilde{D}_{it}, \widetilde{X}_{it}, \widetilde{\epsilon}_{it}) < 0.
\end{aligned}
\tag{1}
$$

All empirical binary choice models, dynamic or static, are special cases of this formulation. The underlying behavioral model that generated the latent variable is dynamic if agents are forward looking *and* either v_{it}^* contains past choices, $\widetilde{D}_{it}$, or unobservables, $\widetilde{\epsilon}_{it}$, that are serially correlated.[7] The underlying model is static (i) if agents are myopic or (ii) if agents are forward looking and there is no link among the past, current and future periods through $\widetilde{D}_{it}$ or serially correlated unobservables.

Researchers may have a number of different, though not necessarily mutually exclusive, goals. They include:

1. Test a prediction of the theory, that is, how an observable variable in v_{it}^* affects d_{it}.

2. Determine the affect of a change in $\widetilde{D}_{it}$ or $\widetilde{X}_{it}$ on choices (either within or outside of the sample variation).

3. Determine the affect of a change in something not in $\widetilde{D}_{it}$ or $\widetilde{X}_{it}$, that is, in something that does not vary in the sample, on choices.

---

[6] As will be seen in the empirical applications we consider, there are a wide range of types of variables that would be included in X. Their common feature is that they are not directly choices of the agent, although they may be affected by prior choices or correlated with choices without being directly affected by them.

[7] By forward looking, we simply mean that agents take into account the effect of their current actions on future welfare. How exactly they form expectations about the impact of those actions and about future preferences and constraints are specific modeling choices.

It is assumed that these statements are ceteris parabus, not only in the sense of conditioning on the other observables, but also in conditioning on the unobservables and their joint conditional (on observables) distribution.[8] Different empirical strategies, for example, structural or non-structural, may be better suited for some of these goals than for others.

III. The Common Empirical Structure of Static and Dynamic Discrete Choice Models

In drawing out the connection between the structure of static and dynamic discrete choice models, it is instructive to consider an explicit example. We take as the prime motivating example one of the oldest and most studied topics in labor economics, the labor force participation of married women.[9] We first illustrate the connection between research goals and empirical strategies in a static framework and then modify the model to allow for dynamics.

A. Married Woman's Labor Force Participation

1. Static Model

Consider the following static model of the labor force participation decision of a married woman. Assume a unitary model in which the couple's utility is given by

$$U_{it} = U(c_{it}, 1 - d_{it}; n_{it}(1 - d_{it}), \kappa_{it}(1 - d_{it}), \epsilon_{it}(1 - d_{it})), \tag{2}$$

where c_{it} is household i's consumption at period t, $d_{it} = 1$ if the wife works and is equal to zero otherwise, n_{it} is the number of young children in the household, and κ_{it} are other observable factors and ϵ_{it} unobservable factors that affect the couple's valuation of the wife's leisure (or home production). In this context, t corresponds to the couple's duration of marriage. The utility function has the usual properties: $\partial U/\partial C > 0, \partial^2 U/\partial C^2 < 0, U(C, 1) > U(C, 0)$.

The wife receives a wage offer of w_{it} in each period t and the husband, who is assumed to work each period, generates income y_{it}. If the wife works, the household incurs a per-child

---

[8]By maintaining the same joint distribution when performing the ceteris parabus change, we are assuming that the change in an observable variable does not induce a change in the joint distribution of unobservables. This assumption is not the same as assuming conditional independence.

[9]The modern approach to this topic began with Mincer (1962).

child-care cost, π, which is assumed to be time-invariant and the same for all households.[10] The household budget constraint is thus

$$c_{it} = y_{it} + w_{it}d_{it} - \pi n_{it}d_{it}. \tag{3}$$

Wage offers are not generally observed for non-workers. It is, thus, necessary to specify a wage offer function to carry out estimation. Let wage offers be generated by

$$w_{it} = w(z_{it}, \eta_{it}), \tag{4}$$

where z_{it} are observable and η_{it} unobservable factors. z_{it} would conventionally contain educational attainment and "potential" work experience (age - education - 6) Unobservable factors that enter the couple's utility function (ϵ_{it}) and unobservable factors that influence the woman's wage offer (η_{it}) are assumed to be mutually serially uncorrelated and to have joint distribution $F_{\epsilon,\eta|y,\kappa,z,n}$.

Substituting (3) into (2) using (4) yields

$$U_{it} = U(y_{it} + w(z_{it}, \eta_{it})d_{it} - \pi n_{it}d_{it}, 1 - d_{it}; n_{it}(1 - d_{it}), \kappa_{it}(1 - d_{it}), \epsilon_{it}(1 - d_{it})), \tag{5}$$

from which we get alternative-specific utilities, U_{it}^1 if the wife works and U_{it}^0 if she does not, namely

$$U_{it}^1 = U(y_{it} + w(z_{it}, \eta_{it}) - \pi n_{it}, 0), U_{it}^0 = U(y_{it}, 1; n_{it}, \kappa_{it}, \epsilon_{it}). \tag{6}$$

The latent variable function, the difference in utilities, $U_{it}^1 - U_{it}^0$, is thus given by

$$v_{it}^* = v^*(y_{it}, z_{it}, n_{it}, \kappa_{it}, \epsilon_{it}, \eta_{it}) \tag{7}$$

The participation decision is determined by the sign of the latent variable: $d_{it} = 1$ if $v_{it}^* \geq 0, d_{it} = 0$ otherwise.

It is useful to distinguish the household's state space, Ω_{it}, consisting of all of the determinants of the household's decision, that is, $y_{it}, z_{it}, n_{it}, \kappa_{it}, \epsilon_{it}, \eta_{it}$, from the part of the state

---

[10]We treat the price of child care as parametric in part to illustrate how alternative approaches to estimation are related to achieving goal 3. A more complete model would allow for a choice among alternative types of child care, for example, of varying qualities, which differ in their price and which may vary over time.

space observable to the researcher, Ω_{it}^-, that is, consisting only of $y_{it}, z_{it}, n_{it}, \kappa_{it}$. Now, define $S(\Omega_{it}^-) = \{\epsilon_{it}, \eta_{it} | v^*(\epsilon_{it}, \eta_{it}; \Omega_{it}^-) > 0\}$ to be the set of values of the unobservables that enter the utility and wage functions that induces a couple with a given observable state space (Ω_{it}^-) to choose $d_{it} = 1$. Then, the probability of choosing $d_{it} = 1$, conditional on Ω_{it}^-, is given by

$$\Pr(d_{it} = 1 | \Omega_{it}^-) = \int_{S(\Omega_{it}^-)} dF_{\epsilon, \eta | y, \kappa, z, n} = G(y_{it}, z_{it}, n_{it}, \kappa_{it}), \tag{8}$$

where $\Pr(d_{it} = 0 | \Omega_{it}^-) = 1 - \Pr(d_{it} = 1 | \Omega_{it}^-)$.

As is clear from (8), $G(y_{it}, z_{it}, n_{it}, \kappa_{it})$ is a composite of three elements of the model: $U(\cdot), w(\cdot), F_{\epsilon, \eta | y, \kappa, z, n}$. These elements comprise the *structure* of the participation model. Structural estimation (S) is concerned with recovering some or all of the structural elements of the model. Non-structural (NS) estimation is concerned with recovering $G(\cdot)$. In principal, each of these estimation approaches can adopt auxiliary assumptions in terms of parametric (P) forms for some or all of the structural elements or for $G(\cdot)$ or be non-parametric (NP). Thus, there are four possible approaches to estimation: NP-NS, P-NS, NP-S and P-S.[11]

We now turn to a discussion about the usefulness of each of these approaches for achieving the three research goals mentioned above. The first research goal, testing the theory, requires that there be at least one testable implication of the model. From (6) and the properties of the utility function, it is clear that an increase in the wage offer increases the utility of working, but has no effect on the utility of not working. Thus, the probability of working for any given agent must be increasing in the wage offer. The second goal, to determine the impact of changing any of the state variables in the model on an individual's participation probability, requires taking the derivative of the participation probability with respect to the state variable of interest. The third goal requires taking the derivative of the participation probability with respect to something that does not vary in the data. That role is played by the unknown child care cost parameter, π. Determining its impact would provide a quantitative assessment of the effect of a child care subsidy on a married woman's labor force participation.[12]

---

[11] In this taxonomy, semi-parametric and semi-structural categories fall into the parametric (P) and structural (S) categories.

[12] As before (see fn. 8), we assume that the change in an observable variable does not induce a change in the joint distribution of unobservables.

Given the structure of the model, to achieve any of these goals, regardless of the estimation approach, it is necessary to adopt an assumption of independence between the unobservable factors affecting preferences and wage offers and the observable factors. Absent such an assumption, variation in the observables, $y_{it}, z_{it}, n_{it}, \kappa_{it}$, either among individuals or over time for a given individual, would cause participation to differ both because of their effect on preferences and/or wage offers and because of their relationship to the unobserved determinants of preferences and/or wage offers through $F_{\epsilon,\eta|y,\kappa,z,n}$. In what follows, we adopt the assumption of full independence, that is, $F_{\epsilon,\eta|y,\kappa,z,n} = F_{\epsilon,\eta}$, so as not to unduly complicate the discussion.

Non-Parametric, Non-Structural: If we make no further assumptions, we can estimate $G(\cdot)$ non-parametrically.

Goal 1: To accomplish the first goal, we need to be able to vary the wage offer independently of other variables that affect participation. To do that, there must be an exclusion restriction, in particular, a variable in z_{it} that is not in κ_{it}. Moreover, determining the sign of the effect of a wage increase on the participation probability requires that knowing the sign of the effect of the variable in z_{it} (not in κ_{it}) on the wage. Of course, if we observed all wage offers, the wage would enter into the latent variable rather than the wage determinants (z_{it} and η_{it}) and the prediction of the theory could be tested directly without an exclusion restriction.

What is the value of such an exercise? Assume that the observation set is large enough that sampling error can be safely ignored and consider the case where all wage offers are observed. Suppose one finds, after non-parametric estimation of the participation probability function, that there is some "small" range of wages over which the probability of participation is declining as the wage increases. Thus, the theory is rejected by the data. Now, suppose we wanted to use the estimated participation probability function to assess the impact of a proportional wage tax on participation. This is easily accomplished by comparing the sample participation probability in the data with the participation probability that comes about by reducing each individual's wage by the tax. Given that the theory is rejected, should we use the participation probability function for this purpose? Should our answer

depend on how large is the range of wages over which the violation occurs? Should we add more κ_{it} variables and retest the model? And, if the model is not rejected after adding those variables, should we then feel comfortable in using it for the tax experiment? If there are no ready answers to these questions in so simple a model, as we believe is the case, then how should we approach them in contexts where the model's predictions are not so transparent and therefore for practical purposes untestable, as is normally the case in DCDP models? Are there other ways to validate models? We leave these as open questions for now, but return to them in the concluding section of the chapter.

Goal 2. Clearly, it is possible, given an estimate of G, to determine the effect on participation of a change in any of the variables within the range of the data. However, one cannot predict the effect of a change in a variable that falls outside of the range of the data.

Goal 3: It is not possible to separately identify G and π. To see that note that because it is πn that enters G, $G_n = \pi G_{(\pi n)}$; knowledge of G_n does not allow one to separately identify $G_{(\pi n)}$ and π. We thus cannot perform the child care subsidy policy experiment.

Parametric, Non-Structural: In this approach, one chooses a functional form for G. For example, one might choose a cumulative standard normal function in which the variables in Ω_{it}^- enter as a single index.

Goal 1: As in the NP-NS approach, because of the partial observability of wage offers, testing the model's prediction still requires an exclusion restriction, that is, a variable in z_{it} that is not in κ_{it}.

Goal 2. It is possible, given an estimate of G, to determine the effect on participation of a change in any of the variables not only within, but also outside, of the range of the data.

Goal 3: As in the NP-NS approach, it is not possible to separately identify π from variation in n_{it} because πn_{it} enters G.

Non-Parametric, Structural: In this approach, one would attempt to separately identify $U(\cdot), w(\cdot), F$ from (8) without imposing auxiliary assumption about those functions. This is clearly infeasible when wages are only observed for those who work. [13]

---

[13]Results from Matzkin (1993) apply to the case where all wage offers are observed (regardless of participation). In that case, aside from normalizations, $w(\cdot), U(\cdot)$ and the joint distribution, F are non-parametrically

Parametric, Structural: Although given our taxonomy, there are many possible variations on which functions to impose parametric assumptions, it is too far removed from the aims of this chapter to work through those possibilities.[14] We consider only the case in which all of the structural elements are parametric. Specifically, the structural elements are specified as follows:

$$U_{it} = c_{it} + \alpha_{it}(1 - d_{it}) \text{ with } \alpha_{it} = \kappa_{it}\beta_\kappa + \beta_n n_{it} + \epsilon_{it}, \tag{9}$$

$$c_{it} = y_{it} + w_{it}d_{it} - \pi n_{it}d_{it}, \tag{10}$$

$$w_{it} = z_{it}\gamma + \eta_{it}, \tag{11}$$

$$f(\epsilon_{it}, \eta_{it})\tilde{}N(0, \Lambda), \tag{12}$$

where $\Lambda = \begin{pmatrix} \sigma_\epsilon^2 & \cdot \\ \sigma_{\epsilon\eta} & \sigma_\eta^2 \end{pmatrix}$.[15] This specification of the model leads to a latent variable function, the difference in utilities, $U_{it}^1 - U_{it}^0$, given by

$$v_{it}^*(z_{it}, n_{it}, \kappa_{it}, \eta_{it}, \epsilon_{it}) = z_{it}\gamma - (\pi + \beta_n)n_{it} - \kappa_{it}\beta_\kappa + \eta_{it} - \epsilon_{it} \tag{13}$$

$$= \xi_{it}^*(\Omega_{it}^-) + \xi_{it},$$

where $\xi_{it} = \eta_{it} - \epsilon_{it}$, $\xi_{it}^*(\Omega_{it}^-) = z_{it}\gamma - (\pi + \beta_n)n_{it} - \kappa_{it}\beta_\kappa$ and Ω_{it}^- now consists of z_{it}, n_{it} and κ_{it}.[16]

The likelihood function, incorporating the wage information for those women who work,

---

identified.

[14] Pagan and Ullah (1999), Chapter 7, provides a good introduction to semi-parametric estimation of discrete choice models

[15] The unconventional assumption of normality for the wage distribution (allowing, as it does, for negative wage offers) is adopted in order to obtain a decision rule that is linear and additive in unobservables. We present a more general formulation in later sections.

[16] As we show below, the additive error (ξ_{it}) is convenient in calculating choice probabilities and is maintained for illustrative purposes. However, as we also show below, the addititve structure is fragile. It is lost, for example, if the wage function takes a semi-log form or if the utility function is non-linear in consumption. Note that the linearity and separability of consumption in the utility function implies that husband's income does not enter v_{it}^* and, thus, does not affect the participation decision.

is

$$L(\theta; \kappa_{it_i}, z_{it_i}, n_{it_i}) \quad = \quad \prod_{i=1}^{I} \Pr(d_{it_i} = 1, w_{it_i} | \ \Omega_{it}^-)^{d_{it_i}} \Pr(d_{it_i} = 0 | \Omega_{it_i}^-)^{1-d_{it_i}} \tag{14}$$

$$= \quad \prod_{i=1}^{I} \Pr(\ \xi_{it_i} \geq -\xi_{it_i}^*(\Omega_{it_i}^-), \eta_{it_i} = w_{it_i} - z_{it_i}\gamma)^{d_{it_i}} \Pr(\xi_{it} < -\xi_{it}^*(\Omega_{it}^-))^{1-d_{it_i}}.$$

The parameters to be estimated include $\beta_\kappa, \beta_n, \gamma, \pi, \sigma_\epsilon^2, \sigma_\eta^2$, and $\sigma_{\epsilon\eta}$.[17] First, it is not possible to separately identify the child care cost, π, from the effect of children on the utility of not working, β_n; only $\pi + \beta_n$ is potentially identified. Joint normality is sufficient to identify the wage parameters, γ and σ_η^2, as well as $(\sigma_\eta^2 - \sigma_{\epsilon\eta})/\sigma_\xi$ (Heckman (1978)). The data on work choices identify γ/σ_ξ and β/σ_ξ. To identify σ_ξ, note that there are three possible types of variables that appear in the likelihood function, variables that appear only in z, that is, only in the wage function, variables that appear only in κ, that is, only in the value of leisure function, and variables that appear in both κ and z. Having identified the parameters of the wage function (the $\gamma's$), the identification of σ_ξ (and thus also $\sigma_{\epsilon\eta}$) requires the existence of at least one variable of the first type, that is, a variable that appears only in the wage equation.[18]

Goal 1: As in the NS approaches, there must be an exclusion restriction, in particular, a variable in z_{it} that is not in κ_{it}.

Goal 2. It is possible to determine the effect on participation of a change in any of the variables within and outside of the range of the data.

Goal 3: As noted, it is possible to identify $\pi + \beta_n$. Suppose then that a policy maker is considering implementing a child care subsidy program, where none had previously existed, in which the couple is provided a subsidy of τ dollars if the wife works when there is a young child in the household. The policy maker would want to know the impact of the program on the labor supply of women and the program's budgetary implications. With such a program, the couple's budget constraint under the child care subsidy program is

$$c_{it} = w_{it}d_{it} + y_{it} - (\pi - \tau)d_{it}n_{it}, \tag{15}$$

---

[17]We call $\Pr(d_{it} = 1, w_{it} | \ \Omega_{it}^-)$ a probability, but it is actually a mixed probability for d and a density for w. Note that the Jacobian of the transformation from the wage density to the wage error density is one.

[18]Given the assumptions of the model, full independence of the joint error distribution with respect to observables is not necessary.

where $(\pi - \tau)$ is the net (of subsidy) cost of child care. With the subsidy, the probability that the woman works is

$$\Pr(d_{it} = 1 | \Omega_{it}^-, \tau) = \Phi \left(\frac{z_{it}\gamma - \kappa_{it}\beta_\kappa - (\beta_n + \pi - \tau)n_{it}}{\sigma_\xi} \right), \tag{16}$$

where Φ is the standard normal cumulative. Given identification of $\beta_n + \pi$ from maximizing the likelihood ((14)), to predict the effect of the policy on participation, that is, the difference in the participation probability when τ is positive and when τ is zero, it is necessary, as seen in (16), to have identified σ_ξ. Government outlays on the program would be equal to the subsidy amount times the number of women with young children who work under the subsidy.

It is important to note that the policy effect is estimated without direct policy variation, i.e., we did not need to observe households in both states of the world, with and without the subsidy program. What was critical for identification was (exogenous) variation in the wage (independent of preferences). Wage variation is important in estimating the policy effect because, in the model, the child care cost is a tax on working that is isomorphic to a tax on the wage. Wage variation, independent of preferences, provides *policy-relevant* variation.

To summarize, testing the prediction that participation rises with the wage offer requires an exclusion restriction regardless of the approach. This requirement arises because of the non-observability of wage offers for those that choose not to work.[19] With regard to the second goal, the parametric approach allows extrapolation outside of the sample range of the variables whereas nonparametric approaches do not. Finally, subject to identification, the P-S approach enables the researcher to perform counterfactual exercises, subsidizing the price of child care in the example, even in the absence of variation in the child care price.[20]

---

[19] If all wage offers were observed, it would be possible to achieve all three goals without imposing parametric assumptions or structure. With respect to the policy counterfactual (goal 3), because of the subsidy acts like a wage tax, the effect of the subsidy can be calculated by comparing participation rates of women with a given wage to women with a wage augmented by πn_{it} (see Ichimura and Taber (2002) and Todd and Wolpin (forthcoming)).

[20] Another reason for adopting the P-S estimation approach is that separating out preferences from opportunities (wage offers) helps to understand important social and economic phenomena, for example, in assessing how much of the difference in labor market outcomes of black and white women is due to differences in preferences and how much to differences in wage opportunities. Such an assessment could be useful in the design of public policies aimed at ameliorating those differences.

2. Dynamic Model

In the previously specified static model, there was no connection between the current participation decision and future utility. One way, among many, to introduce dynamic considerations is through human capital accumulation on the job. In particular, suppose that the woman's wage increases with actual work experience, h, as skills are acquired through learning by doing. To capture that, rewrite (11) as

$$w_{it} = z_{it}\gamma_1 + \gamma_2 h_{it} + \eta_{it}, \tag{17}$$

where $h_{it} = \sum_{\tau=1}^{\tau=t-1} d_{i\tau}$ is work experience at the start of period t. Given this specification, working in any period increases all future wage offers. Work experience, h_{it}, evolves according to

$$h_{it} = h_{i,t-1} + d_{i,t-1} \tag{18}$$

where $h_{i1} = 0$.[21]. Thus, at any period t, the woman may have accumulated up to $t-1$ periods of work experience. We will be more specific about the evolution of the other state space elements when we work through the solution method below. For now, we assume only that their evolution is non-stochastic.

Normally distributed additive shocks: As in the static model, and again for presentation purposes, we assume that the preference shock (ϵ_{it}) and the wife's wage shock (η_{it}) are distributed joint normal. In addition, we assume that they are mutually serially independent and independent of observables, that is, $(f(\epsilon_{it}, \eta_{it}|\epsilon_{it-1}, \eta_{it-1,....}, \epsilon_{i1}, \eta_{i1}) = f(\epsilon_{it}, \eta_{it}))$.

Assume, in this dynamic context, that the couple maximizes the expected present discounted value of remaining lifetime utility at each period starting from an initial period,

---

[21]The assumption that the woman's initial work experience at the time marriage is zero, which is undoubtedly in many cases untrue, is made for ease of exposition. We discuss in a later section the complications introduced by accounting for the fact that work experience is accumulated prior to marriage and varies across women.

$t = 1$, and ending at period T, the assumed terminal decision period.[22,23] Letting $V_t(\Omega_{it})$ be the maximum expected present discounted value of remaining lifetime utility at $t = 1, ..., T$ given the state space and discount factor δ,

$$V_t(\Omega_{it}) = \max_{d_{it}} E\left\{ \sum_{\tau=t}^{\tau=T} \delta^{\tau-t}[U_{i\tau}^1 d_{i\tau} + U_{i\tau}^0(1 - d_{i\tau})]|\Omega_{i\tau} \right\}. \tag{19}$$

The state space at t consists of the same elements as in the static model augmented to include the amount of accumulated work experience, h_{it}.

The value function $(V_t(\Omega_{it}))$ can be written as the maximum over the two alternative-specific value functions, $V_t^k(\Omega_{it})$, $k \in \{0, 1\}$

$$V_t(\Omega_{it}) = \max(V_t^0(\Omega_{it}), V_t^1(\Omega_{it})), \tag{20}$$

each of which obeys the Bellman equation

$$
\begin{aligned}
V_t^k(\Omega_{it}) &= U_{it}^k(\Omega_{it}) + \delta E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}, d_{it} = k] \ \text{ for } t < T, \tag{21} \\
&= U_{iT}^k(\Omega_{iT}) \ \text{ for } t = T.
\end{aligned}
$$

The expectation in (21) is taken over the distribution of the random components of the state space at $t + 1$, $\epsilon_{i,t+1}$ and $\eta_{i,t+1}$, conditional on the state space elements at t.

The latent variable in the dynamic case is the difference in alternative-specific value

---

[22] The finite horizon assumption is immaterial for the points we wish to make. If the current period utility is bounded at all $t = 1, .., \infty$ and the discount factor is less than one, then the solution to the infinite horizon problem can be approximated arbitrarily closely by the solution to a long but finite horizon problem. The essential difference between a finite and infinite horizon model in terms of the predictions about behavior is that in the finite horizon case there are implications for age patterns in behavior.

[23] The terminal period of the model would be at the termnation of the marriage or the retirement of the wife. Accounting for divorce, even taking it to be exogenous, would unduly complicate the model. For illustrative purposes, then, we assume that the wife retires at $T+1$. The value function at $T+1$ is normalized to zero, although a more complete formulation would make the retirement decision of both spouses a choice and would, at the least, specify the determination of post-retirement income through the social security system.

functions, $V_t^1(\Omega_{it}) - V_t^0(\Omega_{it})$, namely[24]

$$v_t^*(\Omega_{it}) = z_{it}\gamma_1 + \gamma_2 h_{it} - \pi n_{it} - \kappa_{it}\beta_\kappa - \epsilon_{it} + \eta_{it} \qquad (22)$$
$$+\delta\left\{[E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}, d_{it} = 1] - [E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}, d_{it} = 0]\right\}$$
$$= \xi_{it}^*(\Omega_{it}^-) + \xi_{it}.^{[25]} \qquad (23)$$

Comparing the latent variable functions in the dynamic (22) and static (13) cases, the only difference is the appearance in the dynamic model of the difference in the future component of the expected value functions under the two alternatives. This observation was a key insight in the development of estimation approaches for DCDP models.

To calculate these alternative-specific value functions, note first that $\Omega_{i,t+1}^-$, the observable part of the state space at $t + 1$, is fully determined by Ω_{it}^- and the choice at t, d_{it}. Thus, one needs to be able to calculate $E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}, d_{it})$ at all values of $\Omega_{i,t+1}^-$ that may be reached from the state space elements at t and a choice at t. A full solution of the dynamic programming problem consists, then, of finding $EV_\tau(\Omega_{i\tau}) = E\max[(V_\tau^0(\Omega_{i\tau}), V_\tau^1(\Omega_{i\tau}))]$ for all values of $\Omega_{i\tau}^-$ at all $\tau = 2, ...T$. We denote this function by $E\max(\Omega_{it}^-)$ or $E\max_t$ for short.

In the finite horizon model we are considering, the solution method is by backwards recursion. However, there are a number of additional details about the model that must first be addressed. Specifically, it is necessary to assume something about how the exogenous observable state variables evolve, that is, $z_{it}, n_{it}, \kappa_{it}$.[26] For ease of presentation, to avoid having to specify the transition processes of the exogenous state variables, we assume that $z_{it} = z_i$ and $\kappa_{it} = \kappa_i$

The number of young children, however, is obviously not constant over the life cycle. But, after the woman reaches the end of her fecund period, the evolution of n_{it} is non-stochastic.[27]

---

[24]Given the lack of separate identification of π and β_n, we set $\beta_n = 0$ to reduce notation.

[26]Because of the linearity and additive separability of consumption in utility, husband's income does not affect the participation decision. We therefore do not need to specify what is known about future husband's income (see below). Again, this assumption is made so that the solution method can be illustrated most effectively.

[27]Later, we introduce stochastic fertility, allowing for the decision model to begin at the time of marriage, when we consider an extension of the model to a multinomial choice setting.

To continue the example, we restrict attention to the woman's post-fecund period. Thus, during that period n_{it} is perfectly foreseen, although the future path of n_{it} at any t depends on the exact ages of the young children in the household at t.[28] Thus, the ages of existing young children at t are elements of the state space at t, Ω_{it}^-.

As seen in (21), to calculate the alternative-specific value functions at period $T-1$ for each element of $\Omega_{i,T-1}^-$, we need to calculate what we have referred to above as $E \max_T$. Using the fact that, under normality, $E(\epsilon_{iT}|\xi_{iT} < -\xi_{iT}^*(\Omega_{iT}^-)) = -\frac{\sigma_{\epsilon\xi}}{\sigma_\xi}\frac{\phi(-\xi_{iT}^*(\Omega_{iT}^-))}{\Phi(-\xi_{iT}^*(\Omega_{iT}^-))}$ and $E(\eta_{iT}|\xi_{iT} \geq -\xi_{iT}^*(\Omega_{iT}^-) = \frac{\sigma_{\eta\xi}}{\sigma_\xi}\frac{\phi(-\xi_{iT}^*(\Omega_{iT}^-))}{1-\Phi(-\xi_{iT}^*(\Omega_{iT}^-))}$, we get

$$E \max{}_T = y_{iT} + (\kappa_i\beta_\kappa)\Phi(-\xi_{iT}^*(\Omega_{iT}^-)) + (z_i\gamma_1 + \gamma_2 h_{iT} - \pi n_{iT})(1 - \Phi(-\xi_{iT}^*(\Omega_{iT}^-)) \quad (24)$$
$$+ \sigma_\xi\phi(-\xi_{iT}^*(\Omega_{iT}^-)).^{29}$$

Note that evaluating this expression requires an integration (the normal cdf) which has no closed form; it thus must be computed numerically. The right hand side of (24) is a function of $y_{iT}, z_i, \kappa_i, n_{iT}$ and h_{iT}.[30] Given a set of model parameters, the $E \max_T$ function takes on a scalar value for each element of its arguments. Noting that $h_{iT} = h_{i,T-1} + d_{i,T-1}$, and being explicit about the elements of $E \max_T$, the alternative-specific value functions at T-1 are (dropping the i subscript for convenience):

$$V_{T-1}^0(\Omega_{T-1}) = y_{T-1} + \kappa\beta_\kappa + \epsilon_{T-1} + \delta E \max(y_T, z, \kappa, n_T, h_{T-1}), \quad (25)$$

$$V_{T-1}^1(\Omega_{T-1}) = y_{T-1} + z\gamma_1 + \gamma_2 h_{T-1} - \pi n_{T-1} + \eta_{T-1} \quad (26)$$
$$+ \delta E \max(y_T, z, \kappa, n_T, h_{T-1} + 1).$$

Thus,

$$v_{T-1}^*(\Omega_{i,T-1}) = z\gamma_1 + \gamma_2 h_{T-1} - \pi n_{T-1} - \kappa\beta_\kappa - \epsilon_{T-1} + \eta_{T-1} \quad (27)$$
$$+ \delta \{E \max(y_T, z, \kappa, n_T, h_{T-1} + 1) - E \max(y_T, z, \kappa, n_T, h_{T-1})\}$$
$$= \xi_{T-1}^*(\Omega_{T-1}^-) + \xi_{T-1}. \quad (28)$$

---

[28] Suppose we define a young child as a child under the age of six (that is, not of school age). Consider a couple who at the start of the woman's infecund period has a 3 year old child and thus for whom $n_{it} = 1$. Then, for that couple, $n_{it+1} = 1, n_{it+2} = 1$ and $n_{it+2} = \cdots = n_{iT} = 0$.

[30] Although n_{iT} would surely be zero at some point, we carry it along to emphasize its perfect foresight property.

As before, because y_T enters both $E\max(y_T, z, \kappa, n_T, h_{T-1}+1)$ and $E\max(y_T, z, \kappa, n_T, h_{T-1})$ additively, it drops out of $\xi^*_{T-1}(\Omega^-_{T-1})$ and thus out of v^*_{T-1}.[31]

To calculate the T-2 alternative-specific value functions, we will need to calculate $E\max_{T-1}$. Following the development for period T-1,

$$
\begin{aligned}
E\max_{T-1} \;=\; & y_{T-1} + (\kappa\beta_\kappa + \delta E\max(y_{T-1}, z, \kappa, n_T, h_{T-1}))\Phi(-\xi^*_{T-1}(\Omega^-_{T-1})) \qquad (29) \\
& + (z\gamma_1 + \gamma_2 h_{T-1} - \pi n_{T-1} + \delta E\max(y_{T-1}, z, \kappa, n_T, h_{T-1}+1))(1 - \Phi(-\xi^*_{T-1}(\Omega^-_{T-1}))) \\[4pt]
& + \sigma_\xi \phi(-\xi^*_{T-1}(\Omega^-_{T-1})).
\end{aligned}
$$

The right hand side of (29) is a function of $y_{T-1}, z, \kappa, n_{T-1}, n_T$ and h_{T-1}. As with $E\max_T$, given a set of model parameters, the $E\max_{T-1}$ function takes on a scalar value for each element of its arguments. Noting that $h_{T-1} = h_{T-2} + d_{T-2}$, the alternative-specific value functions at T-2 and the latent variable function are given by

$$
\begin{aligned}
V^0_{T-2}(\Omega_{T-2}) \;=\; & y_{T-2} + \kappa\beta_\kappa + \epsilon_{T-2} + \delta E\max(y_{T-1}, z, \kappa, n_{T-1}, n_T, h_{T-2}), \qquad (30) \\
V^1_{T-2}(\Omega_{T-2}) \;=\; & y_{T-2} + z\gamma_1 + \gamma_2 h_{T-2} - \pi n_{T-2} + \eta_{T-2} \qquad (31) \\
& + \delta E\max(y_{T-1}, z, \kappa, n_{T-1}, n_T, h_{T-2}+1), \\
v^*_{T-2}(\Omega_{T-2}) \;=\; & z\gamma_1 + \gamma_2 h_{.T-2} - \pi n_{T-2} - \kappa\beta_\kappa - \epsilon_{T-2} + \eta_{T-2} \qquad (32) \\
& + \delta\left\{ E\max(y_{T-1}, z, \kappa, n_{T-1}, n_T, h_{T-2}+1) - E\max(y_{T-1}, z, \kappa, n_{T-1}, n_T, h_{T-2}) \right\} \\
\;=\; & \xi^*_{T-2}(\Omega^-_{T-2}) + \xi_{T-2}. \qquad (33)
\end{aligned}
$$

As at T, y_{T-1} drops out of $\xi^*_{T-2}(\Omega^-_{T-2})$ and thus v^*_{T-2}.

We can continue to solve backwards in this fashion. The full solution of the dynamic programming problem is the set of $E\max_t$ functions for all t from $t = 1, .., T$. These $E\max_t$ functions provide all of the information necessary to calculate the cut-off values, the $\xi^*_t(\Omega^-_t)'s$ that are the inputs into the likelihood function.

Estimation of the dynamic model requires that the researcher have data on work experience, h_{it}. More generally, assume that the researcher has longitudinal data for I married couples and denote by t_{1i} and t_{Li} the first and last periods of data observed for married

---

[31] In solving for the latent variable functions, we could thus set $y_t = 0$ (or any other arbitrary value) for all t.

couple i. Note that t_{1i} need not be the first period of marriage (although it may be, subject to the marriage occurring after the woman's fecund period) and t_{Li} need not be the last (although it may be). Denoting θ as the vector of model parameters, the likelihood function is given by

$$L(\theta; data) = \Pi_{i=1}^{i=I} \Pi_{\tau=t_{1i}}^{\tau=t_{Li}} \Pr(d_{i\tau} = 1, w_{i\tau}|\Omega_{i\tau}^-)^{d_{i\tau}} \Pr(d_{i\tau} = 0|\Omega_{i\tau}^-)^{1-d_{i\tau}}, \tag{34}$$

where $\Pr(d_{i\tau} = 1, w_{i\tau}|\Omega_{i\tau}^-) = \Pr(\ \xi_{i\tau} \geq -\xi_{i\tau}^*(\Omega_{i\tau}^-),\ \eta_{i\tau} = w_{i\tau} - z_{i\tau}\gamma_1 - \gamma_2\,h_{i\tau})$ and $\Pr(d_{i\tau} = 0|\Omega_{i\tau}^-) = 1 - \Pr(\ \xi_{i\tau} \geq -\xi_{i\tau}^*(\Omega_{i\tau}^-).$[32]

Given joint normality of ϵ and η, the likelihood function is analytic, namely

$$L(\theta; data) = \prod_{i=1}^{i=I} \prod_{\tau=t_{1i}}^{\tau=t_{Li}} \left\{ \left[1 - \Phi\left(\frac{-\xi_\tau^*(\Omega_{i\tau}^-) - \rho\frac{\sigma_\xi}{\sigma_\eta}\eta_{i\tau}}{\sigma_\xi(1-\rho^2)^{\frac{1}{2}}} \right) \right] \frac{1}{\sigma_\eta}\phi\left(\frac{\eta_{i\tau}}{\sigma_\eta} \right) \right\}^{d_{i\tau}} \tag{35}$$

$$\times \left\{ \Phi\left(\frac{-\xi_\tau^*(\Omega_{i\tau}^-)}{\sigma_\xi} \right) \right\}^{1-d_{i\tau}},$$

where $\eta_{i\tau} = w_{i\tau} - z_{i\tau}\gamma_1 - \gamma_2\,h_{i\tau}$ and where ρ is the correlation coefficient between ξ and η.[33] Estimation proceeds by iterating between the solution of the dynamic programming problem and the likelihood function for alternative sets of parameters. Maximum likelihood estimates are consistent, asymptotically normal and efficient.

Given the solution of the dynamic programming problem for the cut-off values, the $\xi_{it}^*(\Omega_{it}^-)$'s, the estimation of the dynamic model is in principle no different than the estimation of the static model. However, the dynamic problem introduces an additional parameter, the discount factor, δ, and additional assumptions about how households forecast future unobservables.[34] The practical difference in terms of implementation is the computational effort of having to solve the dynamic programming problem in each iteration on the model parameters in maximizing the likelihood function.

---

[32] If the structure does not yield an additive (composite) error, the latent variable function becomes $v_t^*(\Omega_{it}^-, \eta_{it}, \epsilon_{it})$. Calculating the joint regions of η_{it}, ϵ_{it} that determine the probabilities that enter the likelihood function and that are used to calculate the $E\max(\Omega_{it}^-)$ function must, in that case, be done numerically. We address this more general case below.

[33] As in the static case, the Jacobian of the transformation from the density of the wage offer to the density of η is one.

[34] In the current example, couples are assumed to know the full structure of the model and to use it in forming their forecasts of future wage offers and their future preferences.

Identification of the model parameters requires the same exclusion restriction as in the static case, that is, the appearance of at least one variable in the wage equation that does not affect the value of leisure. Work experience, h_{it}, would serve that role if it does not also enter into the value of leisure (κ). A heuristic argument for the identification of the discount factor can be made by noting that the difference in the future component of the expected value functions under the two alternatives in (22) is in general a non-linear function of the state variables and depends on the same set of parameters as in the static case. Rewriting (22) as

$$v_t^*(\Omega_{it}) = z_i\gamma_1 + \gamma_2 h_{it} - \pi n_{it} - \kappa_i\beta_{\kappa} + \delta W_{t+1}(\Omega_{it}^-) - \epsilon_{it} + \eta_{it}, \tag{36}$$

where $W(\cdot)$ is the difference in the future component of the expected value functions, the non-linearities in W_{t+1} that arise from the distributional and functional form assumptions may be sufficient to identify the discount factor.[35]

As in the static model, identification of the model parameters implies that all three research goals previously laid out can be met. In particular, predictions of the theory are testable, the effects on participation of changes in observables that vary in the sample are estimable and a quantitative assessment of the counterfactual child care subsidy is feasible. The effect of such a subsidy will differ from that in a static model as any effect of the subsidy on the current participation decision will be transmitted to future participation decisions through the change in work experience and thus future wages. If a surprise (permanent) subsidy were introduced at some time t, the effect of the subsidy on participation at t would require that the couple's dynamic programming problem be resolved with the subsidy from t to T and the solution compared to that without the subsidy. A pre-announced subsidy to take effect at t would require that the solution be obtained back to the period of the announcement because, given the dynamics, such a program would have effects on participation starting from the date of the announcement.[36]

---

[35] It is possible that in some models additional parameters might enter W_{t+1}, say through the transition functions of state variables (see below for an example). While the same heuristic argument would apply, its validity would be less apparent.

[36] More generally, if agents have beliefs about future policies (or policy changes), such beliefs should be incorporated into the solution and estimation of the decision model.

Independent additive type-1 extreme value errors: When shocks are additive and come from independent type-1 extreme value distributions, as first noted by Rust (1987), the solution to the dynamic programming problem and the choice probability both have closed forms, that is, they do not require a numerical integration as in the additive normal error case. The cdf of an extreme value random variable u is $\exp(-e^{-\frac{u}{\rho}})$ with mean equal to $\rho\gamma$, where γ is Euler's constant, and variance $\frac{\pi^2 \rho^2}{6}$.

Under the extreme value assumption, it can be shown that for period $t = T$ (dropping the i subscript for convenience),

$$
\Pr(d_T = 1|\Omega_T^-) = \exp\left(\frac{z\gamma_1 + \gamma_2 h_T - \pi n_T - \kappa\beta_\kappa}{\rho}\right)\left(1 + \exp(\frac{z\gamma_1 + \gamma_2 h_T - \pi n_T - \kappa\beta_\kappa}{\rho})\right)^{-1} \tag{37}
$$

$$
E\max{}_T = \rho\left\{\gamma + \log\left[\exp\left(\frac{y_T + z\gamma_1 + \gamma_2 h_T - \pi n_T}{\rho}\right) + \exp\left(\frac{y_T + \kappa\beta_\kappa}{\rho}\right)\right]\right\} \tag{38}
$$

$$
= \rho\left\{\gamma + \frac{y_T + z\gamma_1 + \gamma_2 h_T - \pi n_T}{\rho} - \log(\Pr(d_T = 1|\Omega_T^-)\right\}
$$

and for $t < T$,

$$
\Pr(d_t = 1|\Omega_t^-) = \frac{\exp(\frac{z\gamma_1 + \gamma_2 h_t - \pi n_t - \kappa\beta_\kappa + \delta\{E\max_{t+1}(y_{t+1}, z, \kappa, \tilde{n}_{t+1}, h_t + 1) - E\max_{t+1}(y_{t+1}, z, \kappa, \tilde{n}_{t+1}, h_t)\}}{\rho})}{1 + \exp(\frac{z\gamma_1 + \gamma_2 h_t - \pi n_t - \kappa\beta_\kappa + \delta\{E\max_{t+1}(y_{t+1}, z, \kappa, \tilde{n}_{t+1}, h_t + 1) - E\max_{t+1}(y_{t+1}, z, \kappa, \tilde{n}_{t+1}, h_t)\}}{\rho})} \tag{39}
$$

$$
E\max{}_t = \rho\left\{\gamma + \log\left[\exp\left(\frac{V_t^1(\Omega_t^-)}{\rho}\right) + \exp\left(\frac{V_t^0(\Omega_t^-)}{\rho}\right)\right]\right\} \tag{40}
$$

$$
= \rho\left\{\gamma + \frac{y_t + z\gamma_1 + \gamma_2 h_t - \pi n_t + \delta E\max(y_{t+1}, z, \kappa, \tilde{n}_{t+1}, h_t + 1)}{\rho} - \log(\Pr(d_t = 1|\Omega_{it}^-)\right\}
$$

where $\tilde{n}_{t+1}$ denotes the vector of $n_{t+1}, ..., n_T$ values. The solution, as in the case of normal errors, consists of calculating the $E\max_t$ functions by backwards recursion. As seen, unlike the case of normal errors, the $E\max_t$ functions and the choice probabilities have closed form solutions; their calculation does not require a numerical integration.

The extreme value assumption is, however, somewhat problematic in the labor force participation model as structured. For there to be a closed form solution to the DCDP problem, the scale parameter (ρ), and thus the error variance, must be the same for both the preference shock and the wage shock, a rather strong restriction that is unlikely to hold. The root of the problem is that the participation decision rule depends on the wage shock. Suppose, however, that the participation model was modified so that the decision rule no longer included a wage shock. Such a modification could be accomplished in either of two

ways, either by assuming that the wife's wage offer is not observed at the time that the participation decision is made or that the wage is deterministic (but varies over time and across women due to measurement error). In the former case, the wage shock is integrated out in calculating the expected utility of working. while in the latter there is no wage shock entering the decision problem. Then, by adding an independent type-1 extreme value error to the utility when the wife works, the participation decision rule will depend on the difference in two extreme value taste errors, which leads to the closed form expressions given above.

In either case, there is no longer a selection issue with respect to observed wages. Because the observed wage shock is independent of the participation decision, the wage parameters can be estimated by adding the wage density to the likelihood function for participation and any distributional assumption, such as log normality, can be assumed. In addition, as in the case of normal errors, identification of the wage parameters, along with the exclusion restriction already discussed, implies identification of the rest of the model parameters (including the scale parameter). Thus, the three research goals are achievable. Whether the model assumptions necessary to take advantage of the computational gains from adopting the extreme value distribution are warranted raises the issue how models should be judged and which model is "best," a subject we take up later in the chapter.

Unobserved State Variables: We have already encountered unobserved state variables in the labor force participation model, namely the stochastic elements $(\epsilon_{it}, \eta_{it})$ in Ω_t that affect current choices. However, there may be unobserved state variables that have persistent effects through other mechanisms. Such a situation arises, for example, when the distribution of $(\epsilon_{it}, \eta_{it})$ is not independent of past shocks, that is, when $f(\epsilon_{it}, \eta_{it} | \epsilon_{it-1}, \eta_{it-1,...}, \epsilon_{i1}, \eta_{i1}) \neq f(\epsilon_{it}, \eta_{it})$.

A specific example, commonly adopted in the literature, is when shocks have a permanent-transitory structure. For reasons of tractability, it is often assumed that the permanent component takes on a discrete number of values and follows a joint multinomial distribution.

Specifically,

$$\epsilon_{it} = \sum_{m^h=1}^{M} \sum_{m^w=1}^{M} \lambda_{1m^h m^w} 1(type^h = m^h, type^w = m^w) + \omega_{1it}, \tag{41}$$

$$\eta_{it} = \sum_{m^w=1}^{M} \lambda_{2m} 1(type^w = m^w) + \omega_{2it}. \tag{42}$$

where there are M types each of husbands (h) and wives (w), and thus M^2 couple types and where ω_{1it} and ω_{2it} are joint normal and iid over time.[37] Each wife's type is assumed to occur with probability π_{m^w} and each husband's type with probability π_{m^h}, with $\sum_{m^j=1}^{m^j=M} \pi_{m^j} = 1$ for $j = w, h$. A couple's type is defined by their value of $\lambda_{1m^w m^h}$ where the probability of a couple being of type (m^w, m^h) is given by $\pi_{m^w m^h}$, with $\sum_{m^w=1}^{m^w=M} \sum_{m^h=1}^{m^h=M} \pi_{m^w m^h} = 1$.[38] A couple is assumed to know their own and their spouse's type, so the state space is augmented by the husband's and wife's type. Even though types are not known to the researcher, it is convenient to add them to the state variables in what we previously defined as the observable elements of the state space, Ω_{it}^-. The reason is that, unlike the iid shocks ω_{1it} and ω_{2it} which do not enter the $E\max_t$ functions (they are integrated out), the types do enter the $E\max_t$ functions. The dynamic programming problem must be solved for each couple's type.

The likelihood function must also be modified to account for the fact that the types are unobserved. In particular, letting $L_{(m^w, m^h)}$ be the likelihood function for a type (m^w, m^h) couple, the sample likelihood is the product over individuals of the type probability weighted sum of the type-specific likelihoods, namely

$$\prod_i L^i = \sum_{m^w=1}^{M} \sum_{m^h=1}^{M} \pi_{m^w m^h} L^i_{(m^w, m^h)}. \tag{43}$$

A second example is where the joint errors follows an ARIMA process. To illustrate, suppose that the errors follow a first-order autoregressive process, namely that $\epsilon_{i,t} = \rho_e \epsilon_{i,t-1} + \omega_{1it}$ and $\eta_{it} = \rho_\eta \eta_{i,t-1} + \omega_{2it}$, where ω_{1it} and ω_{2it} are joint normal and iid over time. Consider

---

[37] We ignore the possibility that the husband's type also affects his earnings because, in the model as specified, his earnings has no effect on the participation decision. In a more general specification, one would probably add this source of heterogeneity.

[38] There are obviously restrictions across the husband and wife individual type probabilities and couple type probabilities.

again the alternative-specific value functions at t, explicitly accounting for the evolution of the shocks, namely

$$
\begin{aligned}
V_t^k(\Omega_{it}^-, \epsilon_{it}, \eta_{it}) &= U_{it}^k(\Omega_{it}) + \delta E[V_{t+1}(\Omega_{i,t+1}^-, \epsilon_{it+1}, \eta_{it+1}) | \Omega_{it}^-, \epsilon_{it}, \eta_{it}, d_{it} = k] \qquad (44) \\
&= U_{it}^k(\Omega_{it}) + \delta E[V_{t+1}(\Omega_{i,t+1}^-, \rho_e \epsilon_{it} + \omega_{1it+1}, \rho_\eta \eta_{it} + \omega_{2it+1}) | \Omega_{it}^-, \epsilon_{it}, \eta_{it}, d_{it} = k],
\end{aligned}
$$

where the integration is now taken over the joint distribution of ω_{1it+1} and ω_{2it+1} To calculate the alternative-specific value function at t, it is necessary that the $E\max_{t+1}$ function include not only $\Omega_{i,t+1}^-$, as previously specified, but also the shocks at t, ϵ_{it} and η_{it}. Thus, serial correlation augments the state space that enters the $E\max_t$ functions. The added complication is that these state space elements, unlike those we have so far considered, are continuous variables, an issue we discuss later. The likelihood function is also more complicated to calculate as it requires an integration for each couple of dimension equal to the number of observation periods (and there are two additional parameters, ρ_e and ρ_η).[39]

The existence of unobserved state variables creates also a potentially difficult estimation issue with respect to the treatment of initial conditions (Heckman (1981)). Having restricted the model to the period starting at the time the wife is no longer fecund, by that time most women will have accumulated some work experience, i.e., $h_{t_{1i}}$ will not be zero and will vary in the estimation sample. Our estimation discussion implicitly assumed that the woman's "initial" work experience, that is, work experience at t_{1i}, could be treated as exogenous, that is, as uncorrelated with the stochastic elements of the future participation decisions. When there are unobserved initial state variables, permanent types or serially correlated shocks, this assumption is unlikely to hold.

Although we have not specified the labor force participation model governing decisions prior to this period, to avoid accounting for fertility decisions, it is reasonable to suppose that women who worked more while they were of childbearing ages come from a different type distribution than women who worked less, or, in the case in which there are serially correlated shocks, women with greater work experience during the childbearing period may have experienced shocks (to wages or preferences) that are not uncorrelated with those

---

[39]We could combine the permanent-transitory scheme with the AR(1) scheme by allowing the ω_{1it} and ω_{2it} shocks in (41) and (42) to be AR(1).

that arise after. Put differently, it would seem much more reasonable to assume that the same model governs the participation decision during pre- and post-childbearing ages than to assume that there are two different models in which decisions across those periods are stochastically independent (conditional on observables)

There are several possible solutions to the initial conditions problem. Suppose for the sake of exposition, though unrealistically, that all women begin marriage with zero work experience.[40] At the time of marriage, in the case of permanent unobserved heterogeneity, the couple is assumed to be "endowed" with a given set of preferences. A couple who intrinsically places a low value on the wife's leisure will be more likely to choose to have the wife work and thus accumulate work experience. Such women will have accumulated more work experience upon reaching the end of their childbearing years than women in marriages where the wife's value of leisure is intrinsically greater. Thus, when the end of the childbearing years are reached, there will be a correlation between the accumulated work experience of wives and the preference endowment, or type, of couples.

Suppose that participation decisions during the childbearing years were governed by the same behavioral model (modified to account for fertility) as those during the infecund observation period. In particular, suppose that given a couple's type, all shocks (the $\omega's$ in (41) and (42)) are iid. In that case, work experience can be taken as exogenous conditional on a couple's type. To condition the likelihood (43) on initial experience, we specify a type probability function conditional on work experience at the beginning of the infecund period. Specifically, we would replace $\pi_{m^w m^h}$, taken to be scalar parameters in the likelihood function (43), with the type probability function $\pi_{m^w m^h}(h_{t_{1i}})$, where, as previously defined, t_{1i} is the first (post-childbearing) period observed for couple i.[41]

The type probability function can itself be derived using Bayes' rule starting from the true initial decision period (taken to be the start of marriage in this example). Specifically, denoting the couple's endowment pair (m^w, m^h) as "type" and dropping the i subscript,

---

[40]Alternatively as noted, we could assume, unrealistically as well, that the experience that women have at the start of marriage is exogenous with respect to future participation decisions.

[41]We would also need to include any other initial conditions that affect wage offers ($z_i's$) or preferences ($\kappa_i's$), for example, completed schooling.

because

$$L(d_{t_1-1,\dots},d_2,d_1|type)\Pr(type) = \Pr(type|d_{t_1-1,\dots},d_2,d_1)\Pr(d_{t_1-1,\dots},d_2,d_1) \tag{45}$$

$$= \Pr(type|h_{t_1}) \times \sum_{type} L(d_{t_1-1,\dots},d_2,d_1|type)\Pr(type),\tag{46}$$

the type probability function is

$$\pi_{m^w m^h}(h_{t_{1i}}) = \Pr(type|h_{t_1}) = \frac{L(d_{t_1-1,\dots},d_2,d_1|type)\Pr(type)}{\sum_{type} L(d_{t_1-1,\dots},d_2,d_1|type)\Pr(type)}. \tag{47}$$

Estimating the type probability function $\pi_{m^w m^h}(h_{t_{1i}})$ as a non-parametric function of $h_{t_{1i}}$ provides an "exact" solution (subject to sampling error) to the initial conditions problem, yielding type probabilities for each level of experience that would be the same as those obtained if we had solved and estimated the model back to the true initial period and explicitly used (47). Alternatively, because the type probabilities must also be conditioned on all other exogenous state variables (the z and κ variables), perhaps making non-parametric estimation infeasible, estimating a flexible functional form would provide an "approximate" solution.

If the shocks are serially correlated, work experience at the start of the infecund period is correlated with future choices not only because it affects future wages, but also because of the correlation of stochastic shocks across fecund and infecund periods. In that case, as suggested by Heckman (1981) in a non-structural setting, we would need to have data on exogenous initial conditions at the time of the true initial period (taken here to be the start of marriage), when the labor supply decision process is assumed to begin. Given that, we can specify a density for work experience as a function of those exogenous initial conditions at the start of marriage and incorporate it in the likelihood function.[42]

The Curse of Dimensionality: As we have seen, the solution of the dynamic programming problem required that the $E\max_t$ functions be calculated for each point in the state space.

---

[42]If there is both unobserved permanent heterogeneity and serial correlation, and letting Ω_0 be the exogenous initial conditions at the time of marriage, then in the likelihood function (43), $\pi_{m^w m^h}$ would be replaced with $\pi_{m^w m^h}(h_{t_1}, z, \kappa) \times \Pr(h_{t_1}|\Omega_0, z, \kappa)$. Note that Ω_0 must contain a variable other than z and κ in order to identify the effect of h_{t_1} on a couple's type.

If z and κ take on only a finite number of discrete values (e.g., years of schooling, number of children), as does h_t, the solution method simply involves solving for the $E\max_t$ functions at each point in the state space However, if either z or κ contains a continuous variable (or if the shocks follow an ARIMA process, as already discussed), the dimensionality of the problem is infinite and one obviously cannot solve the dynamic programming problem at every state point. Furthermore, one could imagine making the model more complex in ways that would increase the number of state variables and hence the size of the state space, for example, by letting the vector of taste shifters κ include not just number of children but the number of children in different age ranges. In general, in a finite state space problem, the size of the state space grows exponentially with the number of state variables. This is the so-called curse of dimensionality, first associated with Bellman (1957).

Estimation requires that the dynamic programming problem be solved many times - once for each trial parameter vector that is considered in the search for the maximum of the likelihood function (and perhaps at many nearby parameter vectors, to obtain gradients used in a search algorithm). This means that an actual estimation problem will typically involve solving the DP problem thousands of times. Thus, from a practical perspective, it is necessary that one be able to obtain a solution rather quickly for estimation to be feasible. In practice, there are two main ways to do this. One is just to keep the model simple so that the state space is small. But, this precludes studying many interesting problems in which there are a large set of choices that are likely to be interrelated (for example, choices of fertility, labor supply, schooling, marriage and welfare participation).

A second approach, which a number of researchers have pursued in recent years, is to abandon "exact" solutions to DP problems in favor of approximate solutions that can be obtained with greatly reduced computational time. There are three main approximate solution methods that have been discussed in the literature:[43]

1. Discretization: This approach is applicable when the state space is large due to the presence of continuous state variables. The idea is straightforward: simply discretize

_____

[43]Note that we do not discuss methods like Hotz and Miller (1993) here. They propose a method to circumvent having to obtain a full solution of the DP problem while still obtaining parameter estimates, not a method for solving the DP problem (see below).

the continuous variables and solve for the $E \max_t$ functions only on the grid of discretized values. To implement this method one must either (i) modify the law of motion for the state variables so they stay on the discrete grid (e.g., one might work with a discrete AR(1) process) or (ii) employ a method to interpolate between grid points. Clearly, the finer the discretization, the closer the approximation will be to the exact solution. Discretization does not formally break the curse of dimensionality because the time required to compute an approximate solution still increases exponentially as the number of state variables increases. But it can be an effective way to reduce computation time in a model with a given number of state variables.

2. Approximation and interpolation of the $E \max_t$ functions: This approach was originally proposed by Bellman, Kalaba and Kotkin (1963) and extended to the type of models generally of interest to labor economists by Keane and Wolpin (1994). It is applicable when the state space is large either due the presence of continuous state variables or because there are a large number of discrete state variables (or both). In this approach the $E \max_t$ functions are evaluated at a subset of the state points and some method of interpolation is used to evaluate $E \max_t$ at other values of the state space. This approach requires that the $E \max_t$ interpolating functions be specified parametrically. For example, they might be specified as some regression function in the state space elements or as some other approximating function such as a spline. Using the estimated values of the $E \max_t$ rather than the true values is akin to having a non-linear model with specification error. The degree of approximation error is, however, subject to control. In a Monte Carlo study, Keane and Wolpin (1994) provide evidence on the effect of this approximation error on bias of the estimated model parameters under alternative interpolating functions and numbers of state points. Intuitively, as the subset of the state points that are chosen is enlarged and the dimension of the approximating function is increased, the approximation will converge to the true solution.[44]

As with discretization, the approximation/interpolation method does not formally break the curse of dimensionality, except in special cases. This is because the curse of dimensionality applies to polynomial approximation (see Rust (1996)). As the number of state

---

[44]There is no formal proof of this proposition, though, as noted, Keane and Wolpin (1994) provide Monte Carlo evidence for a particular model that supports the intuition.

variables grows larger, the computation time needed to attain a given accuracy in a polynomial approximation to the Emax function grows exponentially.[45] Despite this, the Keane and Wolpin (1994) approach (as well as some closely related variants) has proven to be useful way to reduce computation time in models with large state spaces, and it has been widely applied in recent years. Rather than describe the method in detail here, we will illustrate the method later in a specific application.

3. Randomization: This approach was developed by Rust (1997). It is applicable when the state space is large due the presence of continuous state variables, but it requires that choice variables be discrete and that state variables be continuous. It also imposes important constraints on how the state variables may evolve over time. Specifically, Rust (1997) shows that solving a random Bellman equation can break the curse of dimensionality in the case of DCDP models in which the state space is continuous and evolves stochastically, conditional on the alternative chosen. Note that because work experience is discrete and evolves deterministically in the labor force participation model presented above, this method does not strictly apply. But, suppose instead that we modelled work experience as a continuous random variable with density function $p(h_{t+1}|h_t, d_t) = p(h_t + jI(d_t = 1) - jI(d_t = 0)|h_t, d_t)$ where j is random variable indicating the extent to which working probabilistically augments work experience or not working depletes effective work experience (due to depreciation of skills). The random Bellman equation (ignoring z and κ), the analog of (20), is in that case given by

$$\widehat{V}_{Mt}(h_t) = \max_{d_t} \left[U_t^{d_t}(h_t) + \frac{\delta}{M} \sum_{m=1}^{M} \widehat{V}_{M,t+1}(h_{t+1,m}|h_t, d_t) p(h_{t+1,m}|h_t, d_t) \right], \qquad (48)$$

where $[h_{t+1,1}, ..., h_{t+1,M}] = [h_1, ..., h_M]$ are M randomly drawn state space elements. The approximate value function $\widehat{V}_{Mt}(h_t)$ converges to $V_t(h_t)$ as $M \to \infty$ at a $\sqrt{M}$ rate. Notice that this is still true if (h_t) is a vector of state variables, regardless of the dimension of the vector. Thus, the curse of dimensionality is broken here, exactly analogously to the way

---

[45]Geweke and Keane (1999) give an example where the curse of dimensionality is broken. This is when the Emax can be expressed as a function of the expected value of each alternative. (That is, these expected values are a sufficient statistic for all the state variables that determine them). The size of this set of variables remains fixed at J, where J is the number of alternatives, even as the state space grows larger.

that simulation breaks the curse of dimensionality in approximation of multivariate integrals (while discretization methods and quadrature do not).[46]

The above approach only delivers a solution for the value functions on the grid $[h_1, ..., h_M]$. But forming a likelihood will typically require calculating value functions at other points. A key point is that $\widehat{V}_{Mt}(h_t)$ is, in Rust's terminology, self-approximating. Suppose we wish to construct the alternative specific value function $\widehat{V}_{M_t}^{d_t}(h_t)$ at a point h_t that is not part of the grid $[h_1, ..., h_M]$. Then we simply form:

$$\widehat{V}_{M_t}^{d_t}(h_t) = U_t^{d_t}(d_t) + \delta \sum_{m=1}^{M} \widehat{V}_{M_t}(h_m) \frac{p(h_m|h_t, d_t)}{\sum_{k=1}^{M} p(h_k|h_t, d_t)}. \qquad (49)$$

Notice that, because any state space element at $t + 1$ can be reached from any element at t with some probability given by $p(\cdot| h_t, d_t)$, the value function at t can be calculated from (49) at any element of the state space at t. In contrast to the methods of approximation described above, the value function does not need to be interpolated using an auxiliary interpolating function.[47] This "self-interpolating" feature of the random Bellman equation is also crucial for breaking the curse of dimensionality (which, as noted above, plagues interpolation methods).

Of course, the fact that the randomization method breaks the curse of dimensionality does not mean it will outperform other methods in specific problems. That the method breaks the curse of dimensionality is a statement about its behavior under the hypothetical scenario of expanding the number of state variables. For any given application with a given number of state variables, it is an empirical question whether a method based on discretization, approximation/interpolation or randomization will produce a more accurate approximation in given computation time.[48] Obviously more work is needed on comparing alternative approaches.[49]

---

[46] Technically this is not quite enough, as convergence must be uniform and not just pointwise.

[47] Because h_t is now an unobserved component of the state space, estimation of $p(h_{t+1}|h_t, d_t)$ must be carried out jointly. This would require a distributional assumption for p and raises issues of the separate identification of p and of the effect of h_t on wages.

[48] This is analogous to the fact that the asymptotic properties of competing estimators (under the hypothetical scenario of increasing sample size) do not reveal which will perform best given finite samples.

[49] Stinebrickner(2000) compares several approximation methods in the context of a DCDP model with serially correlated shocks.

B.   The Multinomial Dynamic Discrete Choice Problem

The structure of the labor force decision problem described above was kept simple to provide an accessible introduction to the DCDP methodology. In this section, we extend that model to allow for:

(i) additional choices;

(ii) non-additive errors;

(ii) general functional forms and distributional assumptions.

The binary choice problem considers two mutually exclusive alternatives, the multinomial problem more than two. The treatment of static multinomial choice problems is standard. The dynamic analog to the static multinomial choice problem is conceptually no different than in the binary case. In terms of its representation, it does no injustice to simply allow the number of mutually exclusive alternatives, and thus the number of alternative-specific value functions in (21), to be greater than two. Analogously, if there are $K > 2$ mutually exclusive alternatives, there will be $K - 1$ latent variable functions (relative to one of the alternatives, arbitrarily chosen). The static multinomial choice problem raises computational issues with respect to the calculation of the likelihood function. Having to solve the dynamic multinomial choice problem, that is, for the $E \max(V_t^0(\Omega_{it}), V_t^1(\Omega_{it}), ..., V_t^K(\Omega_{it}))]$ function that enters the multinomial version of (21) at all values of Ω_{it}^- and at all t, adds significant computational burden.

For concreteness, we consider the extension of DCDP models to the case with multiple discrete alternatives by augmenting the dynamic labor force participation model to include a fertility decision in each period so that the model can be extended to childbearing ages. In addition, to capture the intensive work margin, we allow the couple to choose among four labor force alternatives for the wife. We also drop the assumption that errors are additive and normal. In particular, in the binary model we assumed, rather unconventionally, that the wage has an additive error in levels. The usual specification (based on both human capital theory and on empirical fit) is that the *log* wage has an additive error.[50] Although it is necessary to impose functional form and distributional assumptions to solve and estimate

---

[50] The Ben-Porath (1967) model of human capital accumulation leads to a semi-log form and Heckman and Polachek (1974) show using a Box-Cox transformation that a semi-log form is not rejected by the data.

DCDP models, it is not necessary to do so to describe solution and estimation procedures. We therefore do not impose such assumptions, reflecting the fact that the researcher is essentially unconstrained in the choice of parametric and distributional assumptions (subject to identification considerations).

The following example also illustrates the interplay between model development and data. The development of a model requires that the researcher decide on the choice set, on the structural elements of the model and on the arguments of those structural elements. In an ideal world, a researcher, based on prior knowledge, would choose a model, estimate it and provide a means to validate it. However, in part because there are only a few data sets on which to do independent validations and in part because it is not possible to foresee where models will fail to fit important features of data, the process by which DCDP models are developed and empirically implemented involves a process of iterating among the activities of model specification, estimation and model validation (for example, checking model fit). Any empirical researcher will recognize this procedure regardless of whether the estimation approach is structural or non-structural.

A researcher who wished to study the relationship between fertility and labor supply of married women would likely have in mind some notion of a model, and, in that context, begin by exploring the data. A reasonable first step would be to estimate regressions of participation and fertility as functions of "trial" state variables, interpreted as approximations to the decision rules in a DCDP model.[51] As an example, consider a sample of white married women (in their first marriage) taken from the 1979-2004 rounds of the NLSY79. Ages at marriage range from 18 to 43, with 3/4ths of these first marriages occurring before the age of 27. We adopt, as is common in labor supply models, a discrete decision period to be a year.[52] The participation measure consists of four mutually exclusive and exhaustive alter-

---

[51]Approximations to DCDP model decision rules were first discussed in Heckman (1981) and Wolpin (1984). For an empirical application in the labor economics literature, see Keane and Wolpin (2001).

[52]In theory, the period length should correspond to the frequency of decision-making which, in principle, may differ among choice variables. Like the specification of the model stucture (including assumptions about expectations formation and optimization), the discrete time framework is adopted as an approximation. A continuous time framework would be more general, but would require assumptions about the joint process generating decision times for the choice variables.

natives, working less than 500 hours during a calendar year ($d_{it}^0 = 1$), working between 500 and 1499 hours ($d_{it}^1 = 1$), working between 1500 and 2499 hours ($d_{it}^2 = 1$) and working more than 2500 hours ($d_{it}^3 = 1$).[53] The fertility measure is the dichotomous variable indicating whether or not the woman had a birth during the calendar year. The approximate decision rule for participation is estimated by an ordered probit and the fertility decision rule by a binary probit. The variables included in these approximate decision rules, corresponding to the original taxonomy in section II, are $\widetilde{D}_{it} = \{$total hours worked up to t, hours worked in $t - 1$, whether a child was born in $t - 1$, number of children born between $t - 2$ and $t - 5$, number of children ever born, t (years of marriage up to t)$\}$ and $\widetilde{X}_{ijt} = \{$age of wife, age of spouse, schooling of wife, schooling of spouse$\}$. Consistent with any DCDP model, the same state variables enter the approximate decision rules for participation and for fertility. As seen in table 1, the state variables appear to be related to both decision variables and in reasonable ways.[54]

Suppose the researcher is satisfied that the state variables included in the approximate decision rules should be included in the DCDP model. The researcher, however, has to make a choice as to where in the set of structural relationships the specific state variables should appear: the utility function, the market wage function, the husband's earnings function and/or the budget constraint. The researcher also must decide about whether and where to include unobserved heterogeneity and/or serially correlated errors. Some of these decisions will be governed by computational considerations. Partly because of that and partly to avoid overfitting, researchers tend to begin with parsimonious specifications in terms of the size of the state space. The "final" specification evolves through the iterative process described above.

As an example, let the married couple's per-period utility flow include consumption (c_{it}),

---

[53]To the extent that variations in hours worked within those categories represents differences in the choice of optimal hours, the discretization of hours induces measurement error. In the data, the mean and standard deviation of hours based on the categorization (where the categories are assigned 0, 1000, 2000 and 3000 hours) are almost identical to that based on actual anual hours worked. The standard deviation of hours within the categories is 145, 286, 224 and 429.

[54]We have not, however, in this exploratory stage allowed for serially correlated unobservables either through permanent unobserved heterogeneity or serially correlated shocks.

a per-period disutility from each working alternative and a per-period utility flow from the stock of children (N_{it}). The stock of children includes a newborn, that is a child born at the beginning of period t $(n_{it} = 1)$. Thus,

$$U_{it} = U(c_{it}, d_{it}^1, d_{it}^2, d_{it}^3, N_{it}; \epsilon_{it}^1, \epsilon_{it}^2, \epsilon_{it}^3, \epsilon_{it}^n), \qquad (50)$$

where the $\epsilon_{it}^1, \epsilon_{it}^2, \epsilon_{it}^3$, and ϵ_{it}^n are time-varying preference shocks associated with each of the four choices that are assumed to be mutually serially uncorrelated. Allowing for unobserved heterogeneity, the type specification is (following (41))

$$\epsilon_{it}^j = \sum_{m^h=1}^{M} \sum_{m^w=1}^{M} \lambda_{1m}^j 1(type^h = m^h, type^w = m^w) + \omega_{1it}^j, \ j = 1, 2, 3, n, \qquad (51)$$

where the $\omega^{j\prime}s$ are mutually serially independent shocks.

The household budget constraint incorporates a cost of avoiding a birth (contraceptive costs, b_0, which, for biological reasons, will be a function of the wife's age (her age at marriage, a_0^w, plus the duration of marriage, t) and (child) age-specific monetary costs of supplying children with consumption goods (b_{1k}) and with child care if the woman works $(b_{2k}$ per work hour). Household income is the sum of husband's earnings (y_{it}) and wife's earnings, the product of an hourly wage (w_{it}) and hours worked (1000 hours if $d_{it}^1 = 1$, 2000 hours if d_{it}^2, 3000 hours $d_{it}^3 = 1$). Specifically, the budget constraint is

$$c_{it} = y_{it} + w_{it}(1000d_{it}^1 + 2000d_{it}^2 + 3000d_{it}^3) - b_0(a_0^w + t)(1 - n_{it}) - \sum_{k=1}^{K} b_{1k}N_{kit} \qquad (52)$$

$$- \sum_{k=1}^{K} b_{2k}N_{kit}(1000d_{it}^1 + 2000d_{it}^2 + 3000d_{it}^3)$$

where N_{kit} are the number of children in K different age classes, e.g., 0-1, 2-5, etc.[55] To simplify, we do not allow for uncertainty about births. A couple can choose to have a birth (with probability one) and thus not pay the contraceptive cost or choose not to have a birth (with probability one) and pay the avoidance cost.[56].

---

[55] The constant term in the contraceptive cost function, say b_{00} cannot be separately identified from b_{11}, that is, the goods cost of a newborn (a child age 0-1). Note that $N_{1it} = n_{it}$.

[56] One could instead allow for a choice of whether to contracept or not with pregnancy being an uncertain outocome. We ignore this extension for ease of presentation.

The wife's Ben Porath-Griliches wage offer function depends on her level of human capital, Ψ_{it}, which is assumed to be a function of the wife's completed schooling (S_i^w), assumed fixed after marriage, the wife's work experience, that is, the number of hours worked up to t, E_{it}, and on the number of hours worked in the previous period:

$$\log w_{it} = \sum_{j=1}^{3} \log r^j d_{it}^j + \log \Psi_{it}(S_i^w, E_{it}, d_{it-1}^1, d_{it-1}^2, d_{it-1}^3; \eta_{it}^w), \tag{53}$$

$$\eta_{it}^w = \sum_{m^w=1}^{M} \lambda_{2m^w} 1(type^w = m^w) + \omega_{2it}^w, \tag{54}$$

where the r^j are (assumed to be time-invariant) competitively determined skill rental prices that may differ by hours worked and η_{it}^w is a time varying shock to the wife's human capital following a permanent (discrete type)-transitory scheme.[57] Husband's earnings depends on his human capital according to:

$$\log y_{it} = \log r^h + \log \Psi_{it}^h(S_i^h, a_t^h; \eta_{it}^h), \tag{55}$$

$$\eta_{it}^h = \sum_{m^h=1}^{M} \lambda_{2m^h}^h 1(type^h = m^h) + \omega_{2it}^y. \tag{56}$$

where S_i^h is the husband's schooling and $a_t^h = a_0^h + t$ is his age at t (his age at marriage plus t).[58]

The time-varying state variables, the stock of children (older than one) of different ages, the total stock of children and work experience, evolve according to:

$$N_{2it} = \sum_{j=t-1}^{t-5} n_{ij}; \; N_{3it} = \sum_{j=t-6}^{t-17} n_{ij}; \; N_{it} = N_{it-1} + n_{it}, \tag{57}$$

$$E_{it} = E_{it-1} + 1000 d_{it-1}^1 + 2000 d_{it-1}^2 + 3000 d_{it-1}^3. \tag{58}$$

The state variables in Ω_t^-, augmented to include type, consist of the stock of children (older than one) of different ages, the wife's work experience and previous period work

---

[57] In Ben-Porath's (1967) model of the production of human capital, an individual's wage was given by the product of a human capital per-unit rental price times the individual's human capital stock. Griliches (1976) operationalized the human capital production function as depending on arguments such as schooling, work experience and ability.

[58] Husband's are assumed to work full-time, which implies that, given schooling, age and work experience are isomorphic.

status, the husband's and wife's age at marriage, the husband and wife's schooling levels and the couple's type. The choice set during periods when the wife is fecund, assumed to have a known terminal period (t_m), consists of the four work alternatives plus the decision of whether or not to have a child. There are thus eight mutually exclusive choices, given by $d_{it}^{hn} = \{d_{it}^{00}, d_{it}^{10}, d_{it}^{20}, d_{it}^{30}, d_{it}^{01}, d_{it}^{11}, d_{it}^{21}, d_{it}^{31} : t = 1, ..., t_m - 1\}$, where the first superscript refers to the work choice $(h = \{0, 1, 2, 3\})$ and the second to the fertility choice $(n = \{0, 1\})$.[59] When the wife is no longer fecund, $n_{it} = 0$ and the choice set consists only of the four mutually exclusive alternatives, $d_{it}^{hn} = \{d_{it}^{00}, d_{it}^{10}, d_{it}^{20}, d_{it}^{30} : t = t_m, ..., T\}$.

The objective function of the couple is, as in the binary case, to choose the mutually exclusive alternative at each t that maximizes the remaining expected discounted value of the couple's lifetime utility. Defining U_{it}^{hn} to be the contemporaneous utility flow for the work and fertility choices, the alternative-specific value functions for the multinomial choice problem are

$$
\begin{aligned}
V_t^{hn}(\Omega_{it}) &= U_{it}^{hn}(\Omega_{it}) + \delta E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}^-, d_{it}^{hn}] \text{ for } t < T, \quad (59) \\
&= U_{iT}^{hn}(\Omega_{iT}) \text{ for } t = T,
\end{aligned}
$$

where, letting $\widetilde{V_t^{hn}}$ be the vector of alternative specific value functions relevant at period t,

$$
V_t(\Omega_{it}) = \max(\widetilde{V_t^{hn}}(\Omega_{it})), \quad (60)
$$

and where the expectation in (59) is taken over the joint distribution of the preference and income shocks, $f(\omega_{1t}^1, \omega_{1t}^2, \omega_{1t}^3, \omega_{1t}^n, \omega_{2t}^w, \omega_{2t}^y)$.[60] The $\omega' s$ may have a general contemporaneous correlation structure, but, as noted, are mutually serially independent.

The model is solved by backwards recursion. The solution requires, as in the binary case, that the $E \max_t$ function be calculated at each state point and for all t. In the model as it is now specified, the $E \max_t$ function is a six-variate integral (over the preference shocks, the wife's wage shock and the husband's earnings shock). The state space at t consists of all feasible values of $E_{it}, d_{it-1}^1, d_{it-1}^2, d_{it-1}^3, S^w, S^h, N_{it-1}, N_{kit}$ $(k = 2, 3), a_0^h, a_0^w, type^h, type^w$.

---

[59] For convenience, h is in 1000 hour units.

[60] U_t^{hn} is obtained after substituting for the wife's wage and the husband's earnings in the budget constraint and then substituting for consumption in the utility function.

Notice that all of the state variables are discrete and the dimension of the state space is therefore finite. However, the state space, though finite, is huge. The reason is that to keep track of the number of children in each of the three age groups, it is necessary to keep track of the complete sequence of births. If a woman has say 30 fecund periods, the number of possible birth sequences is $2^{30} = 1,073,700,000$. Even without multiplying by the dimension of the other state variables, full solution of the dynamic programming problem is infeasible, leaving aside the iterative process necessary for estimation.

It is thus necessary to use an approximation method, among those previously discussed, for solving the dynamic programming problem, that is, for solving for the $E\max_t$ functions. As an illustration, we present an interpolation method based on regression. To see how it works, consider first the calculation of the $E\max_T$ for any given state space element. At T the woman is no longer fecund, so we need to calculate

$$E\max{}_T = E_{T-1}\max(U_T^{00}(\widetilde{\omega}), U_T^{10}(\widetilde{\omega}), U_T^{20}(\widetilde{\omega}), U_T^{30}(\widetilde{\omega})), \tag{61}$$

where $\widetilde{\omega}$ is the six-tuple vector of shocks. Although this expression is a six-variate integration, at most four of the shocks actually affect U_T^{hn} for any given h, n choice. Given the lack of a closed form expression, $E\max_T$ must be calculated numerically. A straightforward method is Monte Carlo integration. Letting $\widetilde{\omega}_d$ be the dth random draw, $d = 1, ..., D$, from the joint distribution, $f(\omega_1^1, \omega_1^2, \omega_1^3, \omega_1^N, \omega_2^w, \omega_2^h)$, an estimate of $E\max_T$ at say the kth value of the state space in Ω_T^-, Ω_{Tk}^-, is

$$\widehat{E\max}{}_{Tk} = \frac{1}{D}\sum_{d=1}^{D}\max[U_T^{00}(\widetilde{\omega}_d; \Omega_{Tk}^-), U_T^{10}(\widetilde{\omega}_d; \Omega_{Tk}^-), U_T^{20}(\widetilde{\omega}_d; \Omega_{Tk}^-), U_T^{30}(\widetilde{\omega}_d; \Omega_{Tk}^-)]. \tag{62}$$

Given the infeasibility of calculating $\widehat{E\max}_T$ at all points in the state space, suppose one randomly draws K_T state points (without replacement) and calculates the $\widehat{E\max}_T$ function for those K_T state space elements according to (62). We can treat these K_T values of $\widehat{E\max}_T$ as a vector of dependent variables in an interpolating regression

$$\widehat{E\max}{}_{Tk} = g_T(\Omega_{Tk}^-; \gamma_T) + \varsigma_{Tk}, \tag{63}$$

where γ_T is a time T vector of regression coefficients and $g_T(\cdot; \cdot)$ is a flexible function of state

variables.[61] With this interpolating function in hand, estimates of the $E\max_T$ function can be obtained at any state point in the set Ω_T^-.

Given $\widehat{E\max}_T$, we can similarly calculate V_{T-1}^{hn} at a subset of the state points in Ω_{T-1}^-. Using the D draws from $f(\widetilde{\omega})$, the estimate of $E\max_{T-1}$ at the kth state space element is

$$\widehat{E\max}_{T-1,k} = \frac{1}{D}\sum_{d=1}^{D}\max[V_{T-1}^{00}(\widetilde{\omega_d};\Omega_{T-1,k}^-), V_{T-1}^{10}(\widetilde{\omega_d};\Omega_{T-1,k}^-), V_{T-1}^{20}(\widetilde{\omega_d};\Omega_{T-1,k}^-), V_{T-1}^{30}(\widetilde{\omega_d};\Omega_{T-1,k}^-)],$$
(64)

where V_{T-1}^{hn} is given by (59). Using the $\widehat{E\max}_{T-1,k}$ calculated for K_{T-1} randomly drawn state points from Ω_{T-1}^- as the dependent variables in the interpolating function,

$$\widehat{E\max}_{T-1,k} = g_{T-1}(\Omega_{T-1,k}^-;\gamma_{T-1}) + \varsigma_{T-1},$$
(65)

provides estimated values for the $E\max_{T-1}$ function at any state point in the set Ω_{T-1}^-.[62] Continuing this procedure, we can obtain the interpolating functions for all of the $\widehat{E\max}_t$ functions for all t from t_m (the age at which the woman becomes infertile) through T, that is, $g_T, g_{T-1}, ..., g_{t_m}$.

At $t = t_m - 1$, the choice set now includes the birth of a child. All of the $E\max_t$ functions from $t = 1$ to $t_m - 1$ require numerical integrations over the eight mutually exclusive choices based on the joint error distribution $f(\widetilde{\omega})$. At any t within the fecund period, at the kth state point,

$$\widehat{E\max}_{tk} = \frac{1}{D}\sum_{d=1}^{D}\max[V_t^{00}(\widetilde{\omega_d};\Omega_{tk}^-), V_t^{10}(\widetilde{\omega_d};\Omega_{tk}^-), V_t^{20}(\widetilde{\omega_d};\Omega_{tk}^-), V_t^{30}(\widetilde{\omega_d};\Omega_{tk}^-), \quad (66)$$
$$V_t^{01}(\widetilde{\omega_d};\Omega_{tk}^-), V_t^{11}(\widetilde{\omega_d};\Omega_{tk}^-), U_t^{21}(\widetilde{\omega_d};\Omega_{tk}^-), U_t^{31}(\widetilde{\omega_d};\Omega_{tk}^-)]$$

Again taking K_t random draws from the state space at t, we can generate interpolating

---

[61] Keane and Wolpin (1994) discuss various specifications of the regression funtion.

[62] In the labor force participation model, the total number of potential state points increases in t as feasible work experience and numbers of children increase. A researcher might, as the notation indicates, vary the number of randomly drawn state points with t.

functions:[63]

$$\widehat{E\max}_{tk} = g_t(\Omega_{tk}^-; \gamma_t) + \varsigma_{tk} \quad \text{for } t = 1, .., t_m - 1. \tag{67}$$

In the binary case with additive normal errors, the cut-off values for the participation decision, which were the ingredients for the likelihood function calculation, were analytical. Moreover, although the likelihood function (35) did not have a closed form representation, it required the calculation only of a univariate cumulative normal distribution. In the multinomial choice setting we have described, the set of values of the ω vector determining optimal choices and serving as limits of integration in the probabilities associated with the work alternatives that comprise the likelihood function has no analytical form and the likelihood function requires a multivariate integration.

To accommodate these complications, maximum likelihood estimation of the model uses simulation methods. To describe the procedure, let the set of values of $\widetilde{\omega}_t$ for which the hn^{th} choice is optimal at t be denoted by $S_t^{hn}(\Omega_{it}^-) = \{\omega_{1t}^1, \omega_{1t}^2, \omega_{1t}^3, \omega_{1t}^N, \omega_{2t}^w, \omega_{2t}^y | V_t^{hn} = \max(\widetilde{V_t^{hn}})\}$. Consider the probability that a couple chooses neither to work nor have a child, $h_{it} = 0, n_{it} = 0$, in a fecund period $t < t_m$:

$$\Pr(h_{it} = 0, n_{it} = 0|\Omega_{it}^-) = \int\limits_{S_t^{00}(\Omega_{it}^-)} f(\omega_{1t}^1, \omega_{1t}^2, \omega_{1t}^3, \omega_{1t}^N, \omega_{2t}^w, \omega_{2t}^y) d\omega_{1t}^1 d\omega_{1t}^2 d\omega_{1t}^3 d\omega_{1t}^N d\omega_{2t}^w d\omega_{2t}^y. \tag{68}$$

This integral can be simulated by randomly taking $m = 1, ...M$ draws from the joint distribution of ω, with draws denoted by ω_{mt}, and determining the fraction of times that the value function for that alternative is the largest among all eight feasible alternatives, that is,

$$\widehat{\Pr}(h_{it} = 0, n_{it} = 0|\Omega_{it}^-) = \frac{1}{M} \sum_{m=1}^{M} 1[V_{it}^{00}(\widetilde{\omega_{mt}}) = \max(\widetilde{V_{mt}^{hn}}(\Omega_{it}^-))]. \tag{69}$$

One can similarly form an estimate of the probability for other non-work alternatives, namely for $h_{it} = 0, n_{it} = 1$ for any $t < t_m$ and for $h_{it} = 0$ for any $t_m \leq t \leq T$. Recall that for infecund periods, there are only four alternatives because n_{it} is constrained to be zero.

---

[63]Interpolating functions should be chosen with great care. To avoid ovefitting, it is useful to solve the model at more state points than used in estimating the interpolating function and use the additional points for cross-validation. For example, we might solve the model at 4000 state points, estimate the interpolating function on 2000 points and check the fit, say the R^2, using the other 2000 points.

When the wife works, the relevant probability contains the chosen joint alternative $\{h, n\}$ and the observed wage. For concreteness, consider the case where $h_{it} = 2, n_{it} = 1$. Then the likelihood contribution for an individual who works 2000 hours in period t at a wage of w_{it} is

$$\Pr(h_{it} = 2, n_{it} = 1, w_{it}|\Omega_{it}^-) = \Pr(h_{it} = 2, n_{it} = 1|w_{it}, \Omega_{it}^-) \Pr(w_{it}|\Omega_{it}^-) \tag{70}$$

$$= \Pr(w_{it}|\Omega_{it}^-) \int_{S_t^{21}(\Omega_{it}^-)} dF(\omega_{1t}^1, \omega_{1t}^2, \omega_{1t}^3, \omega_{1t}^N, \omega_{2t}^y|\omega_{2t}^w). \tag{71}$$

For illustrative purposes, suppose that the (log) wage equation is additive in η_{it}^w,

$$\log w_{it} = \sum_{j=1}^3 \log r^j d_{it}^j + \log \Psi_{it}(S_i^w, E_{it}, d_{it-1}^1, d_{it-1}^2, d_{it-1}^3) + \eta_{it}^w, \tag{72}$$

$$= \sum_{j=1}^3 \log r^j d_{it}^j + \log \Psi_{it}(S_i^w, E_{it}, d_{it-1}^1, d_{it-1}^2, d_{it-1}^3) + \sum_{m^w=1}^M \lambda_{2m^w} 1(type^w = m^w) + \omega_{2it}^w$$

and further that $\widetilde{\omega}$ is joint normal.[64] With these assumptions, and denoting the deterministic part of the right hand side of (72) by $\overline{\log w_{it}}$, we can write

$$\Pr(h_{it} = 2|w_{it}, \Omega_{it}^-) \Pr(w_{it}|\Omega_{it}^-) \tag{73}$$

$$= \int_{S_t^{21}(\Omega_{it}^-)} dF(\omega_{1t}^1, \omega_{1t}^2, \omega_{1t}^3, \omega_{1t}^N, \omega_{2t}^y|\omega_{2t}^w = \log w_{it} - \overline{\log w_{it}})$$

$$\times \frac{1}{w_{it}\sigma_{\omega_2^w}} \phi\left(\frac{\log w_{it} - \overline{\log w_{it}}}{\sigma_{\omega_2^w}}\right)$$

where $\frac{1}{w_{it}}$ is the Jacobian of the transformation from the distribution of w to the distribution of ω_2^w. Under these assumptions $f(\omega_{1t}^1, \omega_{1t}^2, \omega_{1t}^3, \omega_{1t}^N, \omega_{2t}^y|\omega_{2t}^w)$ is normal and the frequency simulator for the conditional probability takes the same form as (69) except that ω_{2t}^w is set equal to $\log w_{it} - \sum_{j=1}^3 \log r^j d_{it}^j + \log \Psi_{it} + \sum_{m^w=1}^M \lambda_{2m^w} 1(type^w = m^w)$ and the other five $\omega's$ are drawn from $f(\omega_1^1, \omega_1^2, \omega_1^3, \omega_1^N, \omega_2^y|\omega_2^w)$. Thus, denoting the fixed value of ω_{2t}^w as $\widehat{\omega_{2t}^w}$,

$$\Pr(h_{it} = 2, n_{it} = 1|w_{it}, \Omega_{it}^-) \tag{74}$$

$$= \frac{1}{M} \sum_{m=1}^M 1[V_{it}^{21}(\omega_{m1}^1, \omega_{m1}^2, \omega_{m1}^3, \omega_{m1}^N, \omega_{m2}^y, \widehat{\omega_{mt}^w}) = \max(\widetilde{V_{it}^{hn}}(\omega_{m1}^1, \omega_{m1}^2, \omega_{m1}^3, \omega_{m1}^N, \omega_{m2}^y, \widehat{\omega_{mt}^w})].$$

---

[64]Note that the type-specific parameters, $\lambda's$, are essentially the constant terms in the Ψ production function and cannot be separately identified from the skill rental prices.

Although these frequency simulators converge to the true probabilities as $M \to \infty$, there is a practical problem in implementing this approach. Even for large M, the likelihood is not smooth in the parameters, which precludes the use of derivative methods (e.g., BHHH). This lack of smoothness forces the use of non-derivative methods, which converge more slowly. However, frequency simulators can be smoothed , which makes the likelihood function differentiable and improves the performance of optimization routines. One example is the smoothed logit simulator (McFadden (1989)), namely (in the case we just considered),

$$\Pr(h_{it} = 2, n_{it} = 1 | w_{i\tau}, \Omega_{it}^-) = \frac{1}{M} \sum_{m=1}^{M} \frac{\exp\left[(V_{itm}^{21} - \max(\widetilde{V_{itm}^{hn}}))/\tau\right]}{\sum\limits_{\{h,n\}} \exp\left[(V_{itm}^{hn} - \max((\widetilde{V_{itm}^{hn}}))/\tau\right]} \tag{75}$$

where V_{itm}^{hn} is shorthand for the value functions in (74) and τ is a smoothing parameter. As $\tau \to 0$, the RHS converges to the frequency simulator. The other choice probabilities associated with work alternatives are similarly calculated.

1. Alternative Estimation Approaches

Conceptually, any dynamic programming problem that admits to numerical solution can be estimated. In addition to simulated maximum likelihood, researchers have used various alternative simulation estimation methods, including minimum distance estimation, simulated method of moments and indirect inference. There is nothing in the application of these estimation methods to DCDP models that is special, other than having to iterate between solving the dynamic programming problem and minimizing a statistical objective function.

The main limiting factor in estimating DCDP models is the computational burden associated with the iterative process. It is therefore not surprising that there have been continuing efforts to reduce the computational burden of estimating DCDP models. We briefly review two such methods.

A Bayesian Approach: As has been discussed elsewhere (see Geweke and Keane (2000)), it is difficult to apply the Bayesian approach to inference in DCDP models because the posterior distribution of the model parameters given the data is typically intractably complex. Recently, however, computationally practical Bayesian approaches that rely on Markov Chain

41

Monte Carlo (MCMC) methods have been developed by Imai, Jain and Ching (2009) and Norets (2009). We will discuss the Imai et. al (2009) approach in the stationary case, where it is most effective. Thus, we remove time superscripts from the value functions and denote Ω' as the next period state. We also make the parameter vector θ explicit. Thus, corresponding to equations (20) and the (21), we have

$$V(\Omega_i, \theta) = \max_d (V^0(\Omega_i, \theta), V^1(\Omega_i, \theta)), \tag{76}$$

where

$$V^k(\Omega_i, \theta) = U_i^k(\Omega_i, \theta) + \delta E[V(\Omega_i', \theta) | \Omega_i, d_i = k], ...k = 0, 1. \tag{77}$$

The basic idea is to treat not only the parameters but also the values functions and expected value functions as objects that are to be updated on each iteration of the MCMC algorithm. Hence, we add the superscript (s) to the value functions, the expected value functions and the parameters to denote the values of these objects on iteration (s). We use $\widehat{E^{(s)}}$ to denote the approximation to the expected value and $L(\theta^{(s)})$ to denote the likelihood.

The Imai et. al. (2009) algorithm consists of three steps: the parameter update step (using the Metropolis-Hastings algorithm), the Dynamic Programming step, and the expected value approximation step:

1) The Parameter Updating Step (Metropolis-Hastings algorithm)

First, draw a candidate parameter vector from the proposal density $\theta^{(s)*} \sim q(\theta^{(s)*} | \theta^{(s)})$. Then, evaluate the likelihood conditional on $\theta^{(s)*}$ and conditional on $\theta^{(s)}$. Now, form the acceptance probability

$$P = \min \left\{ \frac{L(\theta^{(s)*}) q(\theta^{(s)} | \theta^{(s)*})}{L(\theta^{(s)}) q(\theta^{(s)*} | \theta^{(s)})}, 1 \right\}. \tag{78}$$

We then accept $\theta^{(s)*}$ with probability P, that is,

$$\theta^{(s+1)} = \left\{ \begin{array}{ll} \theta^{(s)*} & \text{with probability } P \\ \theta^{(s)} .... \text{with probability } 1 - p \end{array} \right\}. \tag{79}$$

2) The Dynamic Programming (or Bellman equation iteration) Step

The following Bellman equation step is nested within the parameter updating step:

$$V^{(s)}(\Omega_i, \theta^{(s)}) = \max_d (V^{0(s)}(\Omega_i, \theta^{(s)}), V^{1(s)}(\Omega_i, \theta^{(s)})), \tag{80}$$

$$V^{k(s)}(\Omega_i, \theta^{(s)}) = U_i^k(\Omega_i, \theta^{(s)}) + \delta\widehat{E^{(s)}}[V(\Omega_i', \theta^{(s)})|\Omega_i, d_i = k], ...k = 0, 1. \tag{81}$$

The difficulty here is in obtaining the expected value function approximation that appears on the right hand side of (81). We describe this next.

3) Expected value approximation step.

The expected value function approximation is computed using information from earlier iterations of the MCMC algorithm. The problem that is that, on iteration (s), they have not, in general, yet calculated the value functions at the specific parameter value $\theta^{(s)}$ that they have drawn on iteration (s). Intuitively, the idea is to approximate the expected value functions at $\theta^{(s)}$ by looking at value functions that were already calculated on earlier iterations of the MCMC algorithm, emphasizing parameter values that are in some sense "close" to $\theta^{(s)}$.

Specifically, the expected value function is approximated as

$$\widehat{E^{(s)}}[V(\Omega_i', \theta^{(s)})|\Omega_i, d_i = k] = \frac{1}{N^{(s)}} \sum_{j=1}^{N^{(s)}} V^{(j)}(\Omega_i', \theta^{(j)})W(\theta^{(j)}, \theta^{(s)}), \tag{82}$$

where $\theta^{(j)}$ denotes a parameter value from an earlier iteration (j) of the MCMC algorithm and $V^{(j)}(\Omega_i', \theta^{(j)})$ is the value function at state point Ω_i' that was calculated on iteration (j).[65] Finally, $W(\theta^{(j)}, \theta^{(s)})$ is a weighting function that formalizes the notion of closeness between $\theta^{(s)}$ and $\theta^{(j)}$. Imai et. al. (2009) use weighting function given by

$$W(\theta^{(j)}, \theta^{(s)}) = \frac{K_h(\theta^{(j)}, \theta)}{\sum_{m=1}^{N^{(s)}} K_h(\theta^{(m)}, \theta)}, \tag{83}$$

where K_h is a kernel with bandwidth h.

Under certain conditions, as the number of iterations grows large, the output of this algorithm generates convergence to the posterior distribution of the parameter vector, as well as convergence to the correct (state and parameter contingent) value functions. One condition is "forgetting." That is, the algorithm will typically be initialized using rather

---

[65] Note that in writing (82) we are implicitly assuming that the state space evolves deterministically, conditional on the current state and current choice. Otherwise (82) would require a double sum, where the inner sum is over states that could potentially be reached from Ω_i given the choice $d_i = k$. Norets (2009) handles the stochastically evolving state space case. See also Ching et. al (2010).

arbitrary initial value functions. Hence, the sum in (82) should be taken using a moving window of more recent iterations so early iterations are dropped. Another key point is that, as one iterates, more lagged values of $\theta^{(j)}$ become available, so more values that are "close" to the current $\theta^{(s)}$ will become available. Hence, the bandwidth in the kernel smoother in (83) should become narrower as one iterates. Note that satisfying both the "forgetting" and "narrowing" conditions simultaneously requires that the "moving window" mentioned earlier must expand as one iterates, but not too quickly. Norets (2009) and Imai, Jain and Ching (2009) derive precise rates.

The Bayesian methods described here are in principle applicable to non-stationary models as well. This should be obvious given that a non-stationary model can always be represented as a stationary model with (enough) age specific variables included in the state space. However, this creates the usual curse of dimensionality, as the state space may expand substantially as a result. Unlike, say, the approximate solution algorithm proposed by Keane and Wolpin (1994), these Bayesian algorithms are not designed (or intended) to be methods for handling extremely large state space problems. Combining the two ideas is a useful avenue for future research.

It is worth noting that no DCDP work that we are aware of has ever reported a distribution of policy simulations that accounts for parameter uncertainty; and, it is also rarely done in non-structural work.[66] The Bayesian approach provides a natural way to do this, and Imai et. al (2009) have produced code that generates such a distribution.

A Non-Full Solution Method: Hotz and Miller (1993) developed a method for the implementing DCDP models that does not involve solving the DP model, that is, calculating the $E \max_t$ functions. HM prove that, for additive errors, the $E \max_t$ functions can be written solely as functions of conditional choice probabilities and state variables for any joint distribution of additive shocks. Although the method does not require that errors be distributed extreme value, the computational advantage of the method is best exploited under that assumption.

---

[66]Of course, providing such a distribution is possible without adopting a Bayesian approach,. although it can be computationally burdensome.

Consider again the binary choice model.[67] From (38), one can see that if we have an estimate of the conditional choice probabilities at all state points, $E \max_T$ can also be calculated at all state points. Denoting the (estimate of the) conditional choice probability by $\widehat{\Pr}(d_{iT} = 1|\Omega_{iT}^-)$,

$$\widehat{E} \max_T = \rho \left\{ \gamma + \frac{y_T + z\gamma_1 + \gamma_2 h_T - \pi n}{\rho} - \log(\widehat{\Pr}(d_{iT} = 1|\Omega_{iT}^-)) \right\}. \qquad (84)$$

Consider now period $T - 1$ and suppose we have an estimate of the conditional choice probabilities, $\widehat{\Pr}(d_{iT-1} = 1|\Omega_{iT-1}^-)$. Then,

$$E \max_{T-1} = \rho \left\{ \gamma + \frac{y_{T-1} + z\gamma_1 + \gamma_2 h_{T-1} - \pi n + \delta \widehat{E} \max_T(h_{T-1} + 1)}{\rho} - \log(\widehat{\Pr}(d_{iT-1} = 1|\Omega_{iT-1}^-)) \right\}, \qquad (85)$$

where, for convenience, we have included only work experience in the $\widehat{E} \max_T$ function. We can continue substituting the estimated conditional choice probabilities in this recursive manner, yielding at any t

$$\widehat{E} \max_t = \rho \left\{ \gamma + \frac{y_t + z\gamma_1 + \gamma_2 h_t - \pi n + \delta \widehat{E} \max_{t+1}(h_t + 1)}{\rho} - \log(\widehat{\Pr}(d_{it} = 1|\Omega_{it}^-)) \right\}. \qquad (86)$$

These $\widehat{E} \max_t$ functions can be used in determining the $\xi_{it}^*(\Omega_{it}^-)$ cut-off values that enter the likelihood function.

As with other approaches, there are limitations. First, the empirical strategy involves estimating the conditional choice probabilities from the data (non-parametrically if the data permit). In the case at hand, the conditional choice probabilities correspond to the proportion of women who work for given values of the state variables (for example, for all levels of work experience). To implement this procedure, one needs estimates of the conditional choice probabilities through the final decision period and for each possible value of the state space. Thus, we need longitudinal data that either extends to the end of the decision period or we need to assume that the conditional choice probabilities can be obtained from synthetic cohorts. This latter method requires an assumption of stationarity, that is, in forecasting the conditional choice probabilities of a 30 year old observed in year t when reaching age 60 in year $t + 30$, it's assumed that the 30 year old would face the same decision-making

---

[67]The main insight in the multinomial setting is the same and the extension is straightforward.

environment (for example, the same wage offer function, etc.) as the 60 year old observed in year t. Most DCDP models in the literature which solve the full dynamic programming problem implicitly make such an assumption as well, though it is not dictated by the method.[68] Moreover, it must also be assumed that there are no state variables observed to the agent but unobserved to us; otherwise, we will not be matching the 30 year olds to the 60 year olds the same unobserved state values.[69] Second, the convenience of using additive extreme value errors brings with it the previously discussed limitations of that assumption. Third, the estimates are not efficient, because the fact that the $\widehat{E}\max_t$ functions themselves contain the parameters in the model structure are not taken into account.

IV.   APPLICATIONS

In this section we review the substantive contributions of the DCDP literature to three main areas of labor economics: (i) labor supply (female and male), (ii) job search and (iii) human capital.

A.   Labor Supply

The literature on dynamic labor supply models can be usefully divided into that on females and males. This is because the two literatures have emphasized different aspects of behavior. A key feature of female labor supply is that a large percentage of women (particularly married women) do not work during significant portions of their life cycle. The central role of the decision of whether or not to work has made the DCDP approach more common in the study of female labor supply than in the literature on males.

The literature on women has also emphasized the relationship between participation and human capital accumulation, while tending to ignore saving. This is no accident, because, as Eckstein and Wolpin (1989) note, it is very difficult computationally to handle participation,

---

[68]Lee and Wolpin (2006, forthcoming) allow for (equilibrium) skill prices to change with calendar time due to technical change.

[69]In recent work, Arcidiacono and Miller (2007) have developed methods for extending the HM approach to allow for unobserved state variables. However, there has as yet been no empirical implementation of that approach to a model as rich as those found in the literature.

human capital and saving simultaneously.[70] The literature has also striven to model how fertility, marriage and participation decisions interact.

In contrast, the literature on males has emphasized the continuous choice of hours of work and savings, with participation usually taken as given. Given an assumption of interior solutions, most papers on dynamics of male labor supply have worked with the first order conditions of agents' optimization problems, rather than using the DCDP approach.[71] Nevertheless, at the end of this section, we review an empirical paper on male labor supply (Imai and Keane (2004)) that adapts the DCDP approach to the case of continuous choices of labor supply and consumption.

1. Female Labor Supply

As we have already noted in the previous discussion, the prevalence of non-participation creates a problem for the analysis of labor supply decisions given that a person's market wage rate is not usually observed for non-participants. The classic paper by Heckman (1974) developed a method for estimating a labor supply function (with continuous hours and non-participation) when wages are only observed for workers. In his framework, the labor supply function is estimated jointly with a wage offer function by maximum likelihood.

The possibility of non-participation raises several additional issues. First, participating in the labor market may entail a fixed time and/or money cost (Cogan (1981)). Second, non-participation may lead to a lack of skill appreciation. Thus, the literature on female labor supply has allowed work experience, as a measure of human capital accumulated on

---

[70]Indeed, to our knowledge the only paper that has attempted to do so is Keane and Wolpin (2001). That paper models the labor supply and human capital investment decisions of young men, who often have low participation rates.

[71]For men, strict application of the DCDP approach would require discretization of hours as an approximation to the choice set. In that case, the parallel to the multinomial choice problem considered above is exact. However, the main insight of the DCDP approach to estimation applies as well to continuous choices and to discrete-continuous choices in which the underlying dynamic programming problem is solved based on first-order conditions or Kuhn-Tucker conditions. That insight was simply the observation that because the continuation value (the E max function) is a deterministic function of state variables, the static model and the dynamic programming model have a common empirical structure.

the job, to affect wage offers (e.g., Weiss and Gronau (1981), Eckstein and Wolpin (1989)). Third, the fact that there is heterogeneity in the extent to which women participate over their lifetimes raises the question of the extent to which that heterogeneity is due to permanent (unobserved) differences in preferences for work or to the influence of past work decisions on participation that arise through transitory taste shocks (Heckman and Willis (1977)). Fourth, nonparticipation implies a potentially central role for marriage and fertility decisions.

2. Mincer's (1962) Life Cycle Model

The earliest paper on labor supply of women to adopt formally a life cycle perspective was Mincer (1962). A married woman's labor supply is based on the permanent income of her husband, as well as her market wage and the couple's tastes for market work, home work and children. Given this framework, the observed variation over the life-cycle in a woman's work hours is merely the result of the allocation of work hours to periods when market wages are high relative the value of home time (i.e., inter-temporal substitution).

Based on this framework, Mincer (1962) hypothesized that a transitory change in husband's income, which has no significant effect on his permanent income, should have no impact on a woman's labor supply. Mincer provided some informal evidence on this hypothesis using data from the 1950 Survey of Consumer Expenditures. Taking 6,766 married white women, he stratified them into 12 groups based on husband's education and age and on the presence of young children. He then subdivided each group into households where the husband worked all year vs. those where the husband had a spell of unemployment. Mincer found that women had a higher participation rate if the husband had experienced an unemployment spell. Based on this evidence that women do respond to transitory changes in husband's income, Mincer concluded that a simple life-cycle model (with perfect foresight and no constraints on borrowing) could not adequately describe the data.

Two points are worth noting. First, Mincer (1962) uses households where the husband works all year as a "control group" for similar households (in terms of education, age and children) where the husband experiences an unemployment spell, with unemployment as the "treatment." Thus, one possible explanation for Mincer's finding is that the treatment and control groups differ in unobserved ways, and that women in the treatment group would have

48

worked more regardless. Second, there are alternative explanations that are consistent with a life-cycle model. For instance, depending on the stochastic process for husband's income, unemployment shocks may induce long lived reductions in earnings. It is also possible that leisure time of the husband and wife are non-separable in utility or that unemployed husbands may contribute to home production and/or child care. In either case, unemployment of the husband may reduce the value of home time for the wife.

3. Non-Full Solution Methods of Estimation

The modern structural literature on the estimation of life cycle models of female labor supply begins with Heckman and MaCurdy (1980).[72] They adopt the utility function

$$U_{it} = \alpha_{it} C_{it}^{\eta} + \beta_{it}(H_{\max} - h_{it})^{\gamma} \qquad 0 < \eta < 1, 0 < \gamma < 1, \tag{87}$$

where C_{it} is household $i's$ consumption at t, h_{it} the wife's hours of work, $H_{\max}$ maximum available hours in the period, and α_{it} and β_{it} are taste-shifters. Leisure is given by $L_{it} = (H_{\max} - h_{it})$. Households have perfect foresight about future preferences and wages. The household maximizes its discounted flow of utility over the finite horizon, $t = 0, .., T$,

$$V_i = \sum_{t=0}^{T} \frac{1}{(1+\rho)^t} U_{it}, \tag{88}$$

where ρ is the household's subjective rate of time preference. The household faces the lifetime budget constraint

$$A_{i0} + \sum_{t=0}^{T} \frac{w_{it} h_{it}}{(1+r)^t} = \sum_{t=0}^{T} \frac{C_{it}}{(1+r)^t}, \tag{89}$$

where A_{i0} is the household's initial assets and r is the (constant) rate of interest. Assuming an interior solution, the first-order conditions for all $t = 0, ..., T$ are

$$\frac{\partial U_{it}}{\partial C_{it}} = \left[\frac{(1+\rho)}{(1+r)}\right]^t \lambda_i, \tag{90}$$

$$\frac{\partial U_{it}}{\partial L_{it}} = \left[\frac{(1+\rho)}{(1+r)}\right]^t w_{it}\lambda_i, \tag{91}$$

where λ_i is the marginal utility of wealth at $t = 0$. Using the utility function specification (87), (91) becomes

$$\gamma\beta_{it} L_{it}^{\gamma-1} = \left[\frac{(1+\rho)}{(1+r)}\right]^t w_{it}\lambda_i. \tag{92}$$

---

[72]Their approach builds on the seminal work of MaCurdy (1981) on the labor supply of men.

Taking logs and rearranging yields the Frisch demand function for leisure,

$$\log L_{it} = \frac{1}{\gamma - 1}\{\log w_{it} + \log \lambda_i + t \log\left[\frac{(1+\rho)}{(1+r)}\right] - \log \beta_{it} - \log \gamma\}. \tag{93}$$

To deal with corner solutions, Heckman and MaCurdy (1980) note that a women will choose not to work if the marginal utility of leisure, evaluated at zero hours of work, exceeds the marginal value of working, that is, if

$$\frac{\partial U_{it}}{\partial L_{it}}\Big|_{L_{it}=H_{\max}} > \left[\frac{(1+\rho)}{(1+r)}\right]^t \lambda_i w_{it}, \quad \text{or} \tag{94}$$

$$\gamma \beta_{it} H_{\max}^{\gamma-1} > \left[\frac{(1+\rho)}{(1+r)}\right]^t \lambda_i w_{it}. \tag{95}$$

Taking logs and rearranging, we can express this participation condition as a reservation wage condition, namely

$$h_{it} > 0 \text{ iff } \log w_{it} > -\log \lambda_{i0} - t\log\left[\frac{(1+\rho)}{(1+r)}\right] + \log \beta_{it} + \log \gamma - (1-\gamma)\log H_{\max}. \tag{96}$$

Notice that if the household has a lower level of lifetime wealth, and hence a higher value of λ_i, the reservation wage is correspondingly reduced.

To obtain an estimable model, Heckman and MaCurdy (1980) assume functional forms for the taste shifter β_{it} and for the wage offer function, namely

$$\log \beta_{it} = X_{it}\phi + \eta_{1i} + \epsilon_{1it}, \tag{97}$$

$$\log w_{it} = Z_{it}\theta + \eta_{2i} + \epsilon_{2it}, \tag{98}$$

where X_{it} and Z_{it} are vectors of observables that affect the taste for leisure and market productivity, η_{1i} and η_{2i} are individual permanent components of the taste for leisure and market productivity and ϵ_{1it} and ϵ_{2it} are respective transitory shocks. Substituting (97) and (98) into (93) and (96), we obtain reduced form equations for (i) leisure conditional on participation and (ii) the participation decision rule:

$$\log L_{it} = f_i + X_{it}\frac{\phi}{\gamma - 1} - Z_{it}\frac{\theta}{\gamma - 1} + \frac{1}{\gamma - 1}\log\left[\frac{(1+\rho)}{(1+r)}\right]t + \frac{\epsilon_{2it} - \epsilon_{1it}}{\gamma - 1}, \tag{99}$$

$$h_{it} > 0 \text{ iff } \frac{\epsilon_{2it} - \epsilon_{1it}}{\gamma - 1} > -f_i - X_{it}\frac{\phi}{\gamma - 1} + Z_{it}\frac{\theta}{\gamma - 1} \tag{100}$$

$$-\frac{1}{\gamma - 1}\log\left[\frac{(1+\rho)}{(1+r)}\right]t + \log H_{\max},$$

50

where $f_i = \frac{1}{\gamma - 1}\{\log \lambda_i + \eta_{2i} - \eta_{1i} - \log \gamma\}$ is an individual-specific fixed effect which subsumes the marginal utility of wealth term λ_i as well as the individual permanent components of tastes for work and productivity.

Under the assumptions of the model (i.e., perfect foresight, no borrowing constraints) this fixed effect is time invariant, capturing everything from periods outside of period t relevant for the woman's labor supply decision at time t. For example, in this model it is not necessary to explicitly include the current or future earnings of a married woman's husband, which is captured through λ_i. In principle, it is not even necessary to control explicitly for whether a woman is married, as the woman's marriage history is also built into λ_i. For instance, a single woman is assumed to anticipate the earnings potential of any husband she will eventually marry. Marriage can only enter the model because it shifts tastes for work, not because it alters lifetime wealth.

To estimate the model Heckman and MaCurdy (1980) assume that the stochastic terms ϵ_{1it} and ϵ_{2it} are jointly normal and serially uncorrelated.[73] The hours and participation equations (99) and (100) are estimated jointly with the wage equation (97) by maximum likelihood. The data consist of 30 to 65 year old continuously married white women from the 1968-75 waves of the PSID. There are 672 women who meet the selection criteria, but to estimate the fixed effects f_i, only women who work at least once can be used, leaving 452.[74]

The variables included in the wage equation Z_{it} are potential experience (i.e., age-education-6) and its square along with the local unemployment rate. Because only time varying covariates can be included due to the presence of the fixed effect in the wage equation, education, for example, is not included. The variables included as taste shifters X_{it} are the total number of children, the number of children less than 6, the wife's age, a measure of the number of hours the husband is unemployed, "other" household income, and an indicator for whether the husband is retired or disabled.[75]

The results of the estimation are mostly standard. Tastes for home time are increasing

---

[73] $H_{\max}$ is set at 8760 hours.

[74] If a woman never works, the likelihood of that event is maximized by setting the fixed effect to $-\infty$. Adjustments for this sample selection made little difference to the estimates.

[75] Although the effect of the wife's age may be interpreted as an estimate of $\log\left[\frac{(1+\rho)}{(1+r)}\right]$, it may also reflect changing preferences for leisure with age.

in the number of children and especially the number less than 6. Both "other" income and the husband's hours of unemployment are statistically insignificant, which Heckman and MaCurdy (1980) interpret as evidence that supports the life-cycle model and that contradicts Mincer (1962). But interestingly, the estimate of γ bumps up against its lower bound of zero. This implies a Frisch elasticity of leisure of $\frac{1}{\gamma-1} = -1$. Converting to a Frisch labor supply elasticity, and noting that mean hours worked in the sample is about 1300, we have that

$$
\begin{aligned}
\frac{\partial \log h_{it}}{\partial \log w_{it}} &= \frac{\partial \log h_{it}}{\partial \log L_{it}} \frac{\partial \log L_{it}}{\partial \log w_{it}} = \frac{L_{it}}{H_{\max} - L_{it}} \frac{1}{1-\gamma} \\
&\approx \frac{L_{it}}{h_{it}} = \frac{7460}{1300} = 5.7
\end{aligned}
\tag{101}
$$

which is certainly a large value.

In a subsequent paper, Heckman and MaCurdy (1982) acknowledged that their choice of functional form had implicitly constrained the elasticity of substitution for leisure, and also for hours, to be large. Specifically, if the Frisch elasticity for leisure is $\frac{1}{\gamma-1}$ and we impose $0 < \gamma < 1$, then the elasticity must range from -1 to $-\infty$. Then, for example, if leisure takes up at least two thirds of available time, (101) implies that the Frisch elasticity of labor supply must be at least 2.

Heckman and MaCurdy (1982) report new results based on an additively separable CRRA utility function,

$$
U_{it} = \alpha_{it} \eta^{-1} C_{it}^{\eta} + \beta_{it} \gamma^{-1} (H_{\max} - h_{it})^{\gamma} \qquad \eta < 1, \gamma < 1
\tag{102}
$$

Adopting (102) does not change anything important in terms of the estimating equations, the only difference being that the constant term $\log \gamma$ drops out of the equation for f_i. But now the constraint on γ is only that it be less than one. In fact, Heckman and MaCurdy (1982) estimate $\gamma = -1.44$, which implies a Frisch elasticity of leisure of $\frac{1}{\gamma-1} = -.041$. Interestingly, this still implies a large value of the Frisch elasticity of labor supply equal to 2.35.

The change in the utility specification has some impact on the other parameter estimates. The impact of children on tastes for work becomes larger. The coefficient on income of other household members becomes quantitatively much larger, but is only significant at the 20% level. Heckman and MaCurdy (1982) interpret this result as being "less favorable toward the permanent income hypothesis." Husband unemployment hours also becomes marginally

significant and negative, implying that husband time at home increases the wife's tastes for work.

Heckman and MaCurdy (1980) conduct a second stage estimation where they regress the fixed effects on various determinants of lifetime wealth. Given estimates of the fixed effects, f_i, and given an estimate of γ and the wage equation fixed effects η_{2i}, we can back out estimates of $\log \lambda_i - \eta_{1i}$. Thus, it is possible to isolate only a composite of the marginal utility of wealth minus the fixed effect in tastes for leisure. It turns out that this composite is reduced by wife's education. We would expect education to increase lifetime wealth (thus reducing λ_i) both by increasing own and potential husband's earnings. But the effect of education on tastes for leisure η_{1i} is an empirical question. The result implies either that education increases taste for leisure, or, if it reduces it, that this effect is outweighed by the income effect.

The Heckman and MaCurdy (1980, 1982) papers, as well as earlier work in a static framework by Heckman (1974), do not accommodate fixed costs of work. Within a static model, Cogan (1981) argued that ignoring fixed costs can lead to severe bias in estimates of female labor supply functions. To see the problem, consider the simple quasi-linear utility function given by

$$
\begin{aligned}
U &= C + \beta \frac{(H_{\max} - h)^{1+\gamma}}{1 + \gamma} \\
&= (wh + Y - F) + \beta \frac{(H_{\max} - h)^{1+\gamma}}{1 + \gamma},
\end{aligned}
\tag{103}
$$

where Y represents non-labor income and F represents fixed costs of working (e.g., child care costs). The equation for optimal hours conditional on working is simply

$$
h^* = H_{\max} - \left(\frac{w}{\beta}\right)^{1/\gamma}.
\tag{104}
$$

In the absence of fixed costs the reservation wage (w_R) is

$$
\begin{aligned}
h^* &> 0 \Rightarrow H_{\max} - \left(\frac{w}{\beta}\right)^{1/\gamma} > 0 \;\; \text{or} \\
w &> \beta H_{\max}^{\gamma} = w_R.
\end{aligned}
\tag{105}
$$

However, as Cogan (1981) points out, it is not appropriate to use marginal conditions to determine the participation decision rule in the presence of fixed costs. Instead, it's necessary

to compare the utilities conditional on working and not working, that is,

$$U(h^*) \;=\; w\left[H_{\max} - \left(\frac{w}{\beta}\right)^{1/\gamma}\right] + Y - F + \frac{\beta}{1+\gamma}\left[\left(\frac{w}{\beta}\right)^{1/\gamma}\right]^{1+\gamma} \qquad \text{and} \qquad (106)$$

$$U(0) \;=\; Y + \frac{\beta}{1+\gamma}H_{\max}^{1+\gamma}.$$

Thus, the decision rule for whether to work (whether $U(h^*) > U(0)$) can be expressed as

$$h > 0 \text{ iff } h^* = \left[H_{\max} - \left(\frac{w}{\beta}\right)^{1/\gamma}\right] > \frac{F}{w} + \frac{1}{w}\frac{\beta}{1+\gamma}\left\{H_{\max}^{1+\gamma} - \left[\left(\frac{w}{\beta}\right)^{1/\gamma}\right]^{1+\gamma}\right\} = h_R > 0.$$
$$(107)$$

It is instructive to compare (105), which simply says that the person begins to work when desired hours are positive with (107), which says a person will begin to work only when optimal hours cross a positive threshold value h_R, which Cogan (1981) refers to as reservation hours. Inspection of the right hand side of the inequality in (107) provides intuition for the threshold value; optimal hours conditional on working must be high enough to cover fixed costs plus an additional term which equals the monetized value of the lost utility from leisure.

Thus, as Cogan (1981) describes, in the presence of fixed costs of work the labor supply function is discontinuous, jumping from zero to the reservation hours level when the reservation wage is reached. The specifications assumed in Heckman (1974) and Heckman and MaCurdy (1980, 1982) are not consistent with such behavior. Another key point is that both costs of working (F) and tastes for work (β) enter the participation equation, while only β enters the labor supply equation. Hence, it is possible that a variable like young children could affect fixed costs of working but not tastes for work, that is, that the presence of young children could affect the participation decision but not labor supply conditional on participating.

To estimate labor supply behavior in the presence of fixed costs, Cogan (1981) proposes to jointly estimate a labor supply function as in (104), a reservation hours function as in (107) and an offer wage function. This is in contrast to Heckman's approach of jointly estimating a labor supply function (104), a participation equation based on marginal conditions as in (105) and an offer wage function.

Cogan (1981) compares both approaches using data on married women aged 30 to 34 taken from the 1967 National Longitudinal Survey of Mature Women. In this sample, 898 wives worked and 939 did not. The labor supply and reservation hours functions both include the wife's education and age, number of young children, and husband's earnings. Cogan estimates that fixed costs are substantial (about 28% of average annual earnings), and that a young child raises fixed costs by about a third. He finds that ignoring fixed costs leads to severe overestimates of labor supply elasticities (conditional on work). Cogan's labor supply function implies a Marshallian elasticity of 0.89 at the mean of the data, compared to 2.45 obtained using the Heckman (1974) approach. The Hicks elasticities are 0.93 vs. 2.64.

However, Cogan also shows that the elasticities are rather meaningless in this context. As he notes, a 10% increase in the offer wage to the average non-working woman in the sample would not induce her to enter the labor market. But a 15% increase would induce her to jump to over 1,300 hours. However, an additional 15% wage increase would "only" induce a further increase of 180 hours (or 13.6%).[76]

An important aspect of Cogan (1981) is that he pays close attention to how the model fits the distribution of hours. This is quite unusual in the static literature, where the focus tends to be on estimating elasticities rather than simulating behavior.[77] Cogan finds that the model without fixed costs cannot explain the fact that few people are observed to work very few hours. Indeed, the model without fixed costs has to predict a large fraction of women working few hours to be able to fit the large fraction of women who do not work. As Cogan describes, this leads to a flattening of the labor supply function, which exaggerates wage elasticities (see Cogan (1981) Figure 2). The model with fixed costs provides a much better fit to the data and does not have this problem.

Kimmel and Kniesner (1998) extend the Heckman and MaCurdy (1980, 1982) analysis to include fixed costs of work. That is, they estimate a labor supply equation analogous to

---

[76]Note that this is still a rather large increase, consistent with a Marshallian elasticity of $13.6/15 = 0.90$

[77]The only exceptions we have come across are van Soest, Woittiez and Kapteyn (1990) and Keane and Moffitt (1998). Both papers note that it is rare to observe people working very low levels of hours (the former paper looking at men, the latter looking at single mothers). Van Soest et al (1990) capture this by building in a job offer distribution where few jobs with low levels of hours are available. Keane and Moffitt (1998) build in actual measures of fixed costs of working (e.g., estimates of child care costs).

(99) jointly with a participation decision rule and an offer wage function, namely

$$\log h_{it} = f_{hi} + e_F \log w_{it} + \alpha_h X_{it} + \epsilon_{hit}, \tag{108}$$

$$\Pr(h_{it} > 0) = \Phi(f_{pi} + \beta \log w_{it} + \alpha_p X_{it}). \tag{109}$$

The first equation is the Frisch labor supply function where the fixed effect f_{hi} captures the marginal utility of initial assets along with any fixed effects in tastes for work. The second equation gives the probability of participation, where Φ is the cumulative standard normal. The fixed effect f_{pi} captures not just the marginal utility of wealth and tastes for work, but also individual heterogeneity in the fixed costs of work.

Following Cogan (1981), the existence of fixed costs breaks the tight link between the parameters in the participation and labor supply equations as we previously saw. Thus, there is no necessary relationship between the parameters e_F and α_h in (108) and the parameters β and α_p in (109). In this framework e_F is the conventional Frisch elasticity of labor supply conditional on employment. But, we can also introduce a Frisch participation elasticity given by

$$e_P = \frac{\partial \log \Pr(h_{it} > 0)}{\partial \log w_{it}} = \beta \frac{\phi(\cdot)}{\Phi(\cdot)}, \tag{110}$$

where ϕ is the standard normal density.

Kimmel and Kniesner (1998) estimate this model using data on 2428 women from the Survey of Income Program Participation (SIPP), 68% of whom are married. The tri-annual interview information was collected in May 1983 to April 1986, giving 9 periods of data. The variables included in X_{it} are marital status, children, education and a quadratic in time. The model is estimated in two stages, where in the first stage predicted wages are constructed for workers and non-workers by estimating the wage equation using Heckman's (1979) two-step procedure. The use of predicted wages serves three purposes: (i) to deal with measurement error, (ii) to fill in missing wages and (iii) to deal with possible endogeneity of wages (which would arise if women with high unobserved tastes for work also tend to have high wages). The variables that appear in the wage equation but not in X_{it} are race and a quadratic in age (potential experience).

The estimates imply a Frisch elasticity of 0.66 for employed women, and a Frisch participation elasticity of 2.39. Average hours of the entire population is given by $\bar{h} = P\bar{h}_e$ where

$\overline{h}_e$ is average hours of the employed and P is the percentage employed. Thus we have that

$$\frac{\partial \log \overline{h}}{\partial \log w} = \frac{\partial \log P}{\partial \log w} + \frac{\partial \log \overline{h}_e}{\partial \log w} = 0.66 + 2.39 = 3.05$$

Thus, the participation elasticity is much larger than the hours elasticity. This result provides some justification for models of female labor supply that focus primarily on the participation decision (see below).

Altug and Miller (1998) extend the life-cycle model of Heckman and MaCurdy (1980, 1982) to include human capital accumulation in the form of learning-by-doing. In addition, they incorporate fixed costs of work, state dependence in tastes for leisure, and aggregate shocks. The first step in Altug and Miller (1998) is to estimate the wage offer function, which takes the form

$$\log w_{it} = \log \omega_t + Z_{it}\gamma + \nu_i + \eta_{it} \tag{111}$$

Here Z_{it} is a vector containing work experience, lagged participation and hours, and other observable determinants of skill, ν_i is a time-invariant skill endowment of person i and ω_t is a skill rental price (determined in equilibrium). In estimation, the ν_i can be treated as individual fixed effects and $\log \omega_t$ as time dummies. A key assumption is that η_{it} reflects only measurement error (and not unobserved variation in skill). Given that assumption, no selection bias problem arises if we estimate (111) by OLS only using periods when women are working, provided we include fixed effects.

Altug and Miller (1998) estimate the wage offer function using PSID data from 1967 to 1985. They require that the women reside in a PSID household for at least 6 consecutive years and that they be employed for at least two years (so that the fixed effects, ν_i, can be estimated). This gives a sample of 2169 women. The estimates imply that labor market experience, particularly recent experience, has a large effect on current wages. For instance, a person who worked the average level of hours for the past four years would have current offer wages about 25% higher than someone who had not worked. Interestingly, the lagged participation coefficients are negative while lagged hours coefficients are positive. The implication is that low levels of hours do not increase human capital: one has to work about 500 to 1000 hours to keep skill from depreciating.

The time dummies in the estimation are estimates of the rental price of skill. The rental

price is estimated to be pro-cyclical, falling in the recession years of 1975 and 1980-1982 and rising in 1977, 1983 and 1985. Average wages among all women in the PSID sample are slightly more pro-cyclical than the estimated rental rates. This suggests a compositional effect whereby people with high $\nu_i's$ tend to enter during booms. This is consistent with the mild pro-cyclical bias in aggregate wage measures for males found by Keane, Moffitt and Runkle (1988).

Altug and Miller (1998) assume a current period utility function given by

$$U_{it} = \alpha_{it}\eta^{-1}C_{it}^{\eta} + d_{it}[U_0(X_{0it}) + U_1(X_{1it}, h_{it}) + \epsilon_{1it}] + (1 - d_{it})\epsilon_{0it} \qquad (112)$$

Here the first term is CRRA in consumption, d_{it} is an indicator for positive hours, $U_0(\cdot)$ captures the fixed cost of work, $U_1(\cdot)$ is the disutility of labor, X_{oit} is a vector of demographic variables that reflect the fixed costs of working, X_{1it} includes X_{0it} along with lagged hours of work that shift tastes for leisure hours and ϵ_{1it} and ϵ_{0it} are stochastic shocks to tastes for the work and non-work options, respectively. These shocks can be interpreted as unobserved variation in the fixed cost of work and the value of home time. Additive separability and the distributional assumptions on ϵ_{1it} and ϵ_{0it} play a key role in the estimation procedure, as discussed below.

As in an earlier paper (Altug and Miller (1990)), it is assumed that markets are complete (that all idiosyncratic shocks are perfectly insurable). Given this assumption and the specification of the utility function, the marginal utility of consumption can be shown to be given by

$$\alpha_{it}C_{it}^{\eta-1} = \lambda_{it} = \zeta_i\lambda_t \quad \text{or} \qquad (113)$$

$$\log C_{it} = \frac{1}{\eta - 1}(\log \zeta_i + \log \lambda_t - \log \alpha_{it}) \qquad (114)$$

As seen in (113), perfect insurance implies that λ_{it} can be decomposed into the product of an individual-specific component ζ_i, reflecting the marginal utility of wealth for individual i, and a time varying component λ_t, reflecting aggregate shocks. A person i with a low ζ_i has a relatively low marginal utility of wealth. But, a person's position in the wealth distribution is constant over time. The only source of uncertainty in the marginal utility of wealth over time are aggregate shocks that cause movements in λ_t.

To obtain an estimable equation, let $\log \alpha_{it} = X_{it}\beta + \epsilon_{cit}$ where X_{it} and ϵ_{cit} are observed and unobserved shifters of tastes for consumption, respectively. The consumption equation, (114), can be estimated by fixed effects (or in first differences), assuming the X_{it} are exogenous. Altug and Miller (1998) include household size, children, age and region in X_{it} and the λ_t are estimated as time dummies. The equation is estimated on data from the PSID, which contains only food consumption. As we would expect, the estimated values of λ_t are high in the recession years of 1975 and 1980-1982.

In the final step, Altug and Miller (1998) estimate the first order condition for hours jointly with a participation condition which allows for fixed costs of work. The first order condition for hours is complex because the marginal utility of leisure is not equated to simply the current wage times the marginal utility of consumption. There is an additional term that arises because working today increases future wages and alters future disutilities from work. We refer to this term as the "expected future return to experience."

Altug and Miller deal with this problem using a version of the Hotz and Miller (1993) estimation algorithm.[78] To outline that procedure, first, given estimates of (111) and (114), they back out estimates of the individual effects ν_i and ζ_i. Second, they use non-parametric regression to estimate the probabilities of participation conditional on the state variables, that is, on the estimated values of ν_i and ζ_i, the work history, and a set of demographics (age, education, marital status, race, children, age and region).[79] Third, they assume the ϵ_{1it} and ϵ_{0it} in (112) are iid extreme value shocks, noting that they are the only source of randomness in the current period payoffs from working vs. not working. As in Hotz and Miller (1993), the value functions at any state can be backed out from the conditional choice probabilities calculated in step 2. This allows one to express the "expected future return to experience" terms as a simple function of the conditional participation probabilities (and their derivatives with respect to h_{it}). In the final estimation step, the parameters left to be

---

[78] See Altug and Miller (1998) equations 6.8 and 6.9, which give the final simple expressions for the labor supply and participation equations. Hotz, Miller, Sanders and Smith (1994) develop a simulation method for implementing the Hotz and Miller (1993) conditional choice probability approach.

[79] Note it is important not to include the aggregate prices λ_t and ω_t in these regressions. Agents are assumed not to know the future realizations of these variables and so cannot condition on them when forming expected future payoffs.

estimated are the those associated with the fixed cost of work $U_0(X_{0it})$ and the disutility of labor $U_1(X_{1it}, h_{it})$.

It is important to understand the restrictions in this approach. There can be no stochastic variation in the marginal utility of leisure, because this additional source of randomness would preclude obtaining simple expressions for the expected future return to experience. Having actual productivity shocks instead of only measurement error in wages would have the same effect. And, consumption and leisure must be separable in utility, so that the stochastic term in tastes for consumption does not influence labor supply decisions. Thus, the extreme value error and additive separability assumptions are crucial.

So far, we have discussed approaches based on estimating the first-order condition for optimal labor supply. An alternative is the "life-cycle consistent" or "two-stage budgeting" approach, where one estimates labor supply equations that condition on the full income allocated to a period (MaCurdy (1983)). Using this approach, Blundell and Walker (1986) estimate a life-cycle consistent model of labor supply behavior of married couples. They use data on couples where both the husband and wife work, and the estimation of the labor supply function is done jointly with a probit equation for whether the wife works (to control for selection into the sample). In sharp contrast to Heckman and MaCurdy (1982) and Kimmel and Kniesner (1998), they obtained an (average) Frisch elasticity of labor supply for women of only 0.033. The Hicks elasticity is 0.009. Based on the figures in their paper, we calculate an income effect of -.206 (at the mean of the data) and a Marshallian elasticity of -0.197.

Blundell, Duncan and Meghir (1998) applied this life-cycle consistent approach to married women from UK Family Expenditure Survey 1978 to 1992. UK tax rates were reduced substantially over the period, and the basic idea of the paper is to exploit this variation to help identify labor supply elasticities. As the authors describe, the decline in rates caused different cohorts to face different paths of tax rates. Relative wages for different education groups also changed markedly over this period.

The idea of the paper can be understood as follows. Imagine we group the data by cohort and education level. That is, for each education/cohort we construct group means of hours and wages in each year. We then subtract group and time means from these quantities.

The key assumption in Blundell, Duncan and Meghir (1998) is that any residual variation in wages after taking out group and time means is exogenous. Their leading example of what might cause such residual variation in wages for a group is tax changes that affect groups differentially. Another source of variation would be exogenous technical change that affects groups differently. The key assumption here is that there are no shifts in labor supply behavior within any of the groups over time (e.g., tastes for leisure can vary by cohort/education level, but not within an education/cohort group over time). They also assume that taking out time means purges both hours and wages for all groups from the influence of aggregate shocks, a seemingly strong assumption as time affects (like the business cycle) may well affect different education/skill groups differently.

The simplest way to think about using the grouped data is to think of regressing the group mean of hours on the group mean of wages, after purging these means of group and time effects. An equivalent approach is to use the individual data and proceed in two steps. In the first step regress after-tax wages on time/group interaction dummies, and get the residuals from this regression. In the second step, regress hours on the after-tax wage, time and group dummies and wage residual. Note that we want the wage coefficient to be identified by wage variation within time/groups. The wage equation residual captures other sources of wage variation, as the first stage wage equation controlled for time/group interactions.[80]

The authors also attempt to deal with possible compositional effects of changes in participation rates on the mean of the error term in the labor supply equation (e.g., a higher wage may induce women with higher tastes for leisure to enter the market) by including an inverse Mills' ratio term that is a function of the group/time participation rate. The labor supply equation that Blundell, Duncan and Meghir (1998) actually estimate has the form

$$ h_{it} = \beta \log w_{it} + \gamma [C_{it} - w_{it}(1 - \tau_{it})h_{it}] + X_{it}\phi + d_g + d_t + \delta_w R_{wit} + \delta_c R_{cit} + M(P_{gt}) + \epsilon_{it}, \quad (115) $$

where τ_{it} is the tax rate, the second term is "virtual" non-labor income, X_{it} is a vector of demographic variables (for example, dummy variables for children in various ages), d_g and

---

[80]An alternative computational approach to taking out group and time means is to regress the group mean of hours on the group mean of wages and a complete set of time and group dummies. Then the wage effect is identified purely from the wage variation not explained by time or group. The advantage of the more involved two-step procedure is that the coefficient on the residual provides a test of exogeneity of wages.

d_t are the group and time dummy variables, R_{wit} and R_{cit} are residuals from the first stage regressions of wages and virtual income on the group and time dummies and $M(P_{gt})$ is the Mills' ratio used to correct for non-participation. The authors estimate this hours function by OLS.

To implement this procedure Blundell, Duncan and Meghir (1998) group the FES data into 2 education groups (legal minimum vs. additional education) and 4 cohorts (people born in 1930-1939, 1940-1949, 1950-1959 and 1960-1969), or 8 groups in total. They include only 20 to 50 year old women with employed husbands. This gives 24,626 women of whom 16,781 work. Note that only workers are used to estimate (115), although the full sample is used to estimate the Mills' ratio. One detail is that 2970 of these women are within a few hours of a kink point in the tax schedule. Blundell et al. choose to drop these women from the data and construct an additional Mills' ratio term to deal with the selection bias this creates. They find that the group/time interactions are highly significant in the wage and virtual income equations.

The estimates imply an uncompensated wage elasticity at the mean of the data of 0.17 and a compensated elasticity of 0.20. In a sensitivity test, the authors report results where, in the first stage, the over-identifying instruments are 5 parameters that describe the tax rules interacted with group dummies. This reduces the number of instruments relative to the case where the group dummies were fully interacted with time dummies. It also means that only variation in wages and virtual income specifically induced by tax changes is used to identify the labor supply elasticities. The estimates give an uncompensated elasticity of 0.18 and an essentially zero income effect. Thus, results are little affected.

4. DCDP Models

The first paper to adopt a full solution approach to modelling female labor supply was Eckstein and Wolpin (1989). The main focus of the paper is on how the decision to work today affects wages and tastes for work in the future. Thus, the paper focuses on three of the four issues central to the female labor supply literature (i) fixed costs of working, (ii) human capital accumulation, and (iii) state dependence in tastes for work. To make estimation feasible (particularly given the 1989 computing technology) Eckstein and Wolpin

(1989) make some key simplifying assumptions. First, they ignore savings and assume a static budget constraint. Second, they ignore the choice of hours of work and treat labor supply as a discrete work/no-work decision.

This set of decisions is notable, as it illustrates well the different paths that the male and female life-cycle labor supply literatures have taken. The life-cycle literature on males has emphasized decisions about hours and savings, which Eckstein and Wolpin (1989) ignore, while in most cases ignoring participation, human capital and state dependence, which they stress. This is not a value judgement on either literature, but simply an observation about what aspects of behavior researchers have found most essential to model in each case. The emphasis on participation, human capital and state dependence explains why the female labor supply literature came to the use of DCDP models several years earlier than the male labor supply literature, as these features are very difficult to handle using Euler equation methods.

A third simplifying assumption that Eckstein and Wolpin (1989) make is that they do not model marriage or fertility. To avoid having to model fertility decisions, the paper looks only at women who were at least 39 years old in 1967 (and hence for the most part past child bearing age). The number of children affects the fixed costs of work, but it is treated as a predetermined variable. Marriage is taken as exogenously given. Including marriage and fertility as additional choice variables would not have been feasible given 1989 technology, but, as we will see, incorporating them as choice variables has been the main thrust of the subsequent literature.

Eckstein and Wolpin (1989) assume a utility function for married woman i at age t given by

$$U_{it} = C_{it} + \alpha_1 p_{it} + \alpha_2 C_{it} p_{it} + \alpha_3 X_{it} p_{it} + \alpha_4 N_{it} p_{it} + \alpha_5 S_i p_{it} \tag{116}$$

where p_{it} is an indicator for labor force participation, X_{it} is work experience (the sum of the lagged p_{it}'s), N_{it} is a vector of numbers of children in various age ranges (0-5 and 6-17) and S_i is the woman's completed schooling. The budget constraint is specified as

$$C_{it} = w_{it} p_{it} + y_t^H - c N_{it} - b p_t \tag{117}$$

where w_{it} the wife's wage (annual earnings) if she works and y_t^H is the annual income of the

63

husband (assumed exogenous).[81] The assumption that utility is linear in consumption has some important consequences. First, substitution of (117) into (116) makes clear that we cannot separately identify the fixed cost of work b and the monetary costs of children c from the disutility of work α_1 and the effect of children on the disutility of work α_4. Thus, b and c are normalized to zero.

The second implication of this specification is that the model will exhibit no income effects on labor supply unless consumption and participation interact in the utility function. If $\alpha_2 = 0$, then husband's income will have no impact on the wife's labor supply. A clear pattern in the data is that women with higher income husbands are less likely to work, which would imply that $\alpha_2 < 0$. Thus, to fit the data, consumption and leisure must be complements in utility, although in general, a negative income effect and consumption/leisure complementarity are conceptually distinct phenomena.

Eckstein and Wolpin (1989) assume a standard log earnings function (linear in schooling, quadratic in work experience) with both a stochastic productivity shock and measurement error. A key point is that there are no shocks to tastes for work, so the only stochastic components in the model are the productivity shocks and measurement error. This simplifies the solution to the dynamic programming problem.[82] The solution takes the form of a sequence of reservation wages (contingent on age, work experience and other state variables). The decision rule for participation is simply to work if the offer wage exceeds the reservation wage, which is a deterministic function of the state. The measurement error accounts for cases where women are observed to make decisions that violate this condition.

Eckstein and Wolpin (1989) estimate the model by maximum likelihood using data on 318 white married women from the NLS Mature Women's cohort. The NLS interviewed them 11 times in the 16 years from 1967 to 1982, making it difficult to construct complete employment histories for all the women. To be in the sample, the women had to have at least four consecutive valid years of data on labor force participation and have a spouse present

---

[81] Annual earnings if the woman works are assumed to equal 2000 times the hourly wage rate, regardless of how many hours the woman actually works. This is necessitated by the 1/0 nature of the work decision.

[82] Eckstein and Wolpin (1989) also assume that husband's earnings is a deterministic function of husband's age, a fixed effect, and a schooling/age interaction. If there were taste shocks or shocks to husband's earnings they would have to be integrated out in solving the DP problem.

in every interview from 1967 to 1982. The data set contained 3020 total observations, 53% of which were for working years. The discount factor is fixed at 0.952.

An interesting aspect of the estimates is that they show substantial selection bias in OLS wage equation estimates. The OLS schooling coefficient is 0.08, while the model estimate (which corrects for selection) is 0.05. The experience profile is initially less steep but also less strongly concave than implied by OLS. The estimates also imply that 85% of observed wage variation is measurement error.[83]

With regard to the utility function estimates, Eckstein and Wolpin (1989) find that children (especially young children) negatively effect tastes for work, as expected. The impact of state dependence is imprecisely estimated, but it implies that experience reduces tastes for work. Schooling reduces tastes for work as well. However, both taste effects are clearly outweighed by the positive effects of experience and schooling on wage offers.

Eckstein and Wolpin (1989) find that $\alpha_2 < 0$; thus, as expected, husband income reduces the wife's participation rate. To quantify the size of the income effect, they consider a woman at age 39 with 15 years of work experience, 12 years of schooling, no children and a husband with \$10,000 in annual earnings (which is close to the mean in the data). The baseline prediction of the model is that she will work 5.9 years out of the 21 years through age 59, or 28% of the time. If husband's earnings increase 50% the model predicts her participation rate will drop by half, to 14%. So the elasticity of the participation rate with respect to non-labor income is roughly 1.0. Converting this to an income effect, and noting that the mean wage in the data is \$2.27 dollars per hour and work is assumed to be 2000 hours per year, we obtain an income effect of -0.45.

Unfortunately, Eckstein and Wolpin (1989) do not report a simulation of how an exogenous change in the wage rate (an increase in the intercept, the skill rental price, of the log wage function) would affect labor supply. However, as schooling is exogenous, and the effect of schooling on tastes for work is quantitatively small, we can approximate this using the

---

[83]Note that the measurement error in wages cannot be estimated using wage data alone. But joint estimation of a wage equation and a labor supply model does allow measurement error to estimated, as true wage variation affects behavior while measurement error does not. Of course, any estimate of the extent of measurement error so obtained will be contingent on the behavioral model.

estimated schooling coefficient. Consider, the same representative woman described above, and assume her education level is increased from 12 to 16. An extra 4 years of schooling raises the wage rate roughly 22% at the mean of the data. The model predicts that this will cause her participation rate from age 39 to 59 to increase by 108%. Thus, the implied (uncompensated) elasticity of the participation rate with respect to the wage is roughly 5.0.

Finally, Eckstein and Wolpin (1989) report a detailed description of how the model fits labor force participation rates, conditional on 28 experience and age cells. In general, the model provides a very good fit to the data. As we noted earlier, there are very few papers in the static labor literature, or the literature on dynamic models based on first order conditions, that examine model fit. In contrast, the careful examination of model fit in the DCDP literature has become standard practice. The focus of the former literature is on estimation of parameters or elasticities, while the focus of the DCDP literature is on model simulations under baseline vs. counterfactual scenarios. It is only natural to compare the simulated baseline data to the actual data. Keane and Wolpin (RED, 2009) argue that it ought to be the industry standard to assess model fit in all econometric models (including static models, nonstructural models, etc.).

The next paper in the DCDP literature on female labor supply did not appear until van der Klaauw (1996), which extends Eckstein and Wolpin (1989) to include marriage as a choice. Women have up to 4 options in each period, given by the cross product of work and marriage choices. Another extension is that van der Klaauw (1996) models decisions starting from when a woman has left school (rather than age 39, as in Eckstein and Wolpin), which may be as young as 14. Obviously then, he cannot treat fertility as given. Thus, van der Klaauw (1996) models the arrival of children as a stochastic process, where arrival probabilities depend upon the state variables (i.e., marital status, education, age and race). This is a common practice in DCDP modelling – that is, to take variables that one believes are endogenous, but which one does not wish to model explicitly as a choice (either for computational reasons or because they are not the main focus of the analysis), and treat them as being generated by a stochastic process that depends on the other state variables.[84]

---

[84]This method is not necessarily more parsimonious than modeling a variable as a choice, trading off an additional choice variable (whether to have a child in this case with the corresponding utility and cost

The model is in many ways similar to Eckstein and Wolpin (1989), again incorporating a static budget constraint and a utility function that is linear in consumption. Van der Klaauw specifies the utility function, conditional on the participation (p_t) and marriage choice (m_t), as

$$U_{pm,t} = \alpha_1 m_t + (\alpha_2 + \alpha_3 m_t)p_t + (\beta_1 + \beta_2 p_t + \beta_3 m_t)C_{pm,t} + \epsilon_{pm,t} \qquad (118)$$

Consumption is interacted with participation, as in Eckstein and Wolpin (1989), which enables the model to explain why women work less if they have high income husbands. Tastes for marriage (α_1) are allowed to depend on demographics, children and lagged marriage. Marriage, m_{it}, is also interacted with consumption, C_{it}, thus letting marriage shift the marginal utility of consumption. The effects of demographics, children and lagged participation on tastes for work are captured by letting α_2 and α_3 depend on these variables. There is a separate taste shock for each of the mutually exclusive choices, $\epsilon_{mp,t} = (\epsilon_{11,t}, \epsilon_{10,t}, \epsilon_{01,t}, \epsilon_{00,t})$.

Recall that in Eckstein and Wolpin's (1989) model a woman received utility from total household consumption. Here, a woman is assumed to consume her own income plus a fraction of the husband's income (which depends on her work status), so she receives utility from private consumption. A single women has a probability each year of receiving a marriage offer. The potential husband is characterized by his mean wage, which depends on the woman's characteristics (reflecting marriage market equilibrium) and a transitory wage draw.

It is worth noting that this is a search model of marriage only in a trivial sense. There is no match-specific component to the marriage. That is, a husband does not come with a permanent component to his earnings level, which could make him a "good draw" given the woman's demographics. Nor is there any permanent component to the utility level he provides. Thus, the woman has no reason to decline a marriage offer in the hope of a better offer. Her only reason for systematic delay is that mean husband income is found to be increasing in the woman's potential experience, and thus, her age. This setup substantially reduces the computational burden of estimation, as there is no "husband type" variable that must be included in the state space. But at the same time, the model is not informative

---

parameters) against additional parameters governng the stochastic outcome (the probability of having a child). A limitation of this method is that it does not allow for effects of contemporaneous shocks, for example a high wage draw for the female, on the probability of having a child.

about the effect of permanent differences in husband income on the wife's labor supply, as all permanent differences are a deterministic function of the wife's own characteristics.

The woman's own wage offer function includes standard covariates, such as education, a quadratic in experience, race, age and region. It also includes a lagged participation indicator, which allows recent work experience to be relatively more important. An unusual aspect of the specification, however, is that it is specified in levels, with an additive error. This is also true of the husband's wage function. The reason for adopting this specification is that, when these functions are substituted into the budget constraint to obtain the choice-specific consumption level and this in turn is substituted into the utility function, each of the 4 alternatives turns out to have an additive error that consists of the relevant $\epsilon_{mp,t}$, plus a function of the female and male wage equations errors.

From a computational point of view, what enables handling the additional complexity of making marriage a choice is the assumption that these four additive choice-specific error terms, say $e_{mp,t}$ for $p = 0, 1$, $m = 0, 1$, are assumed to be distributed iid extreme value. As we have previously discussed, this assumption leads to a closed form solutions for the DP problem and for the likelihood function. As also noted, the cost of making the extreme value assumption is that (i) it is contrary to the evidence suggesting that wage errors are approximately log normal and (ii) it assumes that shocks are contemporaneously uncorrelated. This latter assumption is very strong given that the four composite errors contain common error components; for example, $e_{01,t}$ and $e_{11,t}$ have husband income shocks in common.[85]

The model is estimated on PSID data from 1968 to 1985. The sample includes 548 females aged 12 to 19 in 1968 (29 to 36 in 1985), so that complete work and marital histories can be constructed (avoiding the initial conditions problem that would arise for women who were older in 1968). The terminal period is set at age 45 to reduce computational burden. It is assumed that $p_t = 1$ if the woman worked at least 775 hours in a year, but, as in Eckstein and Wolpin (1989), the work choice is assumed to entail 2000 hours of work regardless of actual hours. An approximation is necessary by the binary nature of the work decision.

The model is estimated in stages. In the first stage, the "reduced form" model (with the

---

[85] An alternative approach would be to assume the four errors follow a generalized extreme value distribution (see Arcidiacono (2008)).

woman's and the husband wage equations substituted into (119) is estimated. In the second stage, the wage equations are estimated using employment and marriage decision rules from the reduced from model to implement a selection correction. In the third stage, a minimum distance estimator (see Chamberlain (1984)) is used to recover the structural parameters.

The estimates of the wage equations are a bit difficult to compare to prior literature as they are in levels. For instance, they imply that a year of schooling raises a woman's earnings by $1,379 per year. As mean earnings in the data are $13,698 per year, this is roughly 10% at the mean of the data. A year of schooling also raises potential husband's earnings by $1,266 per year (vs. a mean of $19,800) or 6.4%. This suggests that an important part of the return to schooling for women comes through the marriage market.[86] The utility function estimates imply that children reduce the utility from participation while lagged work increases the utility from participation.

Van der Klaauw (1996) presents a substantial amount of evidence on the fit of the model, showing that it provides a good fit to the proportion of women who are working and married conditional on years since leaving school, to marriage rates by age, and to the hazard functions for marriage and divorce. It also provides a good fit to the proportion of women making each of the 4 marital status/work choices conditional on work experience and age.

Van der Klaauw (1996) then uses the model to simulate the impact of exogenous $1,000 increases in annual offer wages and husband offer wages. The $1,000 wage increase leads to a 26% (i.e., 2.5 year) increase in work experience by age 35. As this is a 7.3% wage increase, this implies an uncompensated labor supply elasticity of roughly 3.6. It is notable, however, that this elasticity is not comparable to a conventional Marshallian elasticity that holds all else fixed. In particular, the wage increase causes a 1 year increase in average years to first marriage, and a 1.3 year decrease in average total years of marriage. The reduction in marriage is part of what induces the increase in labor supply.[87]

---

[86]The estimates imply that a married woman who works receives 34% of husband income. Unfortunately, the share if she does not work is not identified. As can be seen from (118), if a married woman does not work her utility from consumption is $\beta_1 + \beta_3$ times her share of husband income. Only this product is identified in the model.

[87]Van der Klaauw (1996) simulates that a $1000 (or 5%) increase in husband offer wages would reduce average duration to first marriage by 1 year, increase average years of marriage (by age 35) by 2.3 years,

The next significant paper in the DCDP literature on female labor supply is Francesconi (2002), which extends Eckstein and Wolpin (1989) by making fertility a choice and allowing for both full- and part-time work. Thus, women have 6 choices in each annual period (after age 40 only the 3 work options are available). Francesconi (2002) also allows full and part-time experience to have separate effects on wage offers.[88] Thus, the model has three endogenous state variables: number of children, and part-time and full-time experience.

Marriage is taken to be exogenous and the model begins when a woman first gets married and ends at age 65. Women are assumed to make decisions based on the expected value of husband's income. As in van der Klaauw (1996), the husband's mean income is purely a function of the woman's characteristics (i.e., age at marriage, education, education/age of marriage interactions, age). As in Eckstein and Wolpin (1989) women receive utility from total consumption of the household, net of fixed costs of work and costs of children. There is again a static budget constraint, with utility linear in consumption. Utility for woman i at age t, conditional on their part-time and full-time work and fertility choices ($p_t \in (0,1)$, $f_t \in (0,1)$, $n_t \in (0,1)$), is given by

$$U_{it} = C_{it} + \alpha_1 p_{it} + \alpha_2 f_{it} + (\alpha_3 + \epsilon_{it}^n)N_{it} + \alpha_4 N_{it}^2 + (\beta_1 p_{it} + \beta_2 f_{it} + \beta_3 n_{it})C_{it} + (\beta_4 p_{it} + \beta_5 f_{it})n_{it}$$
(119)

The tastes for part and full-time work, α_1 and α_2, are allowed to be a function of the stock of children, N_{it}, work experience and schooling. Tastes for children vary stochastically over time, as captured by ϵ_{it}^n. Consumption is interacted with all the choice variables in order to allow husband's income to affect work and fertility decisions. Work and fertility decisions are interacted, which enables the model to capture the fact that women have lower participation rates during years that they have newborn children.

The stochastic terms in the model are the errors in the full and part-time log wage equations and the shock to tastes for children. There are no additional taste shocks. The

---

and reduce average years of work by 2.6 years, or 27%. These are very large income effects, but they are not comparable to standard income effect measures, as they refer to changes in husband offer wages as opposed to changes in actual husband wages (or changes in some other type of non-labor income). Furthermore, it is not clear how much credence we can give to these figures since, as noted earlier, all permanent differences in husband income in the model are generated by differences in the wife's own characteristics

[88]There are separate part- and full-time wage functions.

errors are assumed to be distributed as joint normal. Thus, as in Eckstein and Wolpin (1989), it is necessary to assume wages are measured with error into account for observations where women are observed to work at wages that are less than the reservation wage. Given that the model contains 6 choices and three s terms the evaluation of the Emax function integrals is difficult. Thus, Francesconi (2002) uses a simulation method like that proposed in Keane and Wolpin (1994) to evaluate the Emax functions. However, the state space is small enough that he can simulate the Emax function at every point in the state space (there is no need to interpolate between points). The three dimensional choice probability integrals are also simulated.

A point worth stressing is that Francesconi (2002) assumes that only the number of children, and not their ages, enters the state space. If children of different ages had different effects on labor supply, as we have previously noted, the size of the state space would grow astronomically. Francesconi can accommodate that newborns have a different effect on labor supply than older children, because newborns are treated as a current choice variable, and they do not enter the state (as they are no longer newborns in the next period). But allowing, e.g., the number of children aged 1 to 5 to have a different effect than the number of children aged 6-17, would greatly increase in complexity.

Francesconi (2002) also follows van der Klaauw (1996) in limiting the size of the state space by assuming husband's mean income is purely a function of the woman's characteristics. Thus, husband-specific characteristics (e.g., a husband skill endowment) need not be included in the state space. Further, it is assumed that husband's earnings are realized only after the wife's labor supply and fertility decisions are made. As a result, the effect of husband's income on the wife's behavior can only be identified to the extent that there are exclusion restrictions, such that certain characteristics of the wife enter the model only through their affect only the husband's wage. In fact, the husband wage function includes the wife's age, age at marriage and education/age of marriage interactions, and all of these variables are excluded from the wife's wage function and from her taste parameters.

Finally, Francesconi (2002) also extends earlier DCDP models of female labor supply by following the procedure in Keane and Wolpin (1997) to allow for unobserved heterogeneity. Specifically, he allows for three discrete types of women in terms of their skill endowments

(the intercepts in the offer wage functions) and in tastes for children (α_3 and α_4).

The model is estimated on a sample of 765 white women from the NLS Young Women Survey who were interviewed 16 times over the 24 years from 1968 to 1991. To be included in the sample the woman must be at least 19 and be continuously married to the same spouse during the sample period.[89] Part-time is defined as 500 to 1,500 hours and full-time is defined as 1,500+ hours. The discount factor is fixed at 0.952. In contrast to the multi-step procedure in van der Klaauw (1996), the decision rules and wage offer functions are estimated jointly. There are separate wage offer functions for part-time and full-time work.

The estimates of the wage function imply that a year of schooling raises the full-time offer wage by 8.4% and the part-time offer wage by 7.6%, estimates that are intermediate between the Eckstein and Wolpin (1989) and van der Klaauw (1996) results. Full-time experience has a larger positive effect on full-time offer wages than part-time experience. Effects of experience on part-time offer wages are generally much smaller. Measurement error accounts for about 63% of the variance of observed wages. Evaluated at the mean of the data, an extra year of school raises mean husband wages by 11%. This is consistent with the finding of van der Klaauw (1996) that a large part of the return to schooling for women comes through the labor market rather than the marriage market. The interaction terms between consumption and work and fertility $(\beta_1, \beta_2, \beta_3)$ are all negative, which generates negative income effects on both labor supply and fertility. In addition, individuals of the type with a high skill endowment have relatively low tastes for children.

Francesconi (2002) reports results indicating that the model provides a good fit to all 6 annual choice options up to 24 years after marriage, which corresponds to age 47 on average (the last observed age in the NLSY79 data he analyzed). He also fits a static model (i.e., a model with the discount factor set to 0) and finds that it too provides a good fit to the in-sample data. But the models differ dramatically in their out-of-sample predictions. The static model predicts that women's labor supply will increase sharply after about age 47 and into the 60's. The DCDP model implies that work will stay flat and then drop slowly in the 60's. The latter prediction is much closer to what is observed in CPS data, which covers

---

[89]This is a sub-sample of a group of 1,783 women who were married at least once during the period (the larger sample including women who leave a partner during the sample period).

adult women of all ages.[90] The static model explains low participation rates as resulting from the presence of children; when children leave the household, participation rates rise sharply. In the dynamic model, the return to human capital investment, that is of working, falls as one approaches the terminal period, which counteracts the effect of children leaving.

Finally, Francesconi (2002) conducts a number of simulations of how permanent changes in wages would affect labor supply. For example, consider an average woman with 2 years of full-time work experience at the time of marriage. The baseline model simulation shows that she will work for 6.8 out of the 11 years from age 30 to 40. An increase in the log wage function intercept (which represents the rental rice of skill) would increase offer wages at the mean of the data by roughly 10.5%, and it would increase full-time work by roughly 60%. This implies an elasticity of labor supply with respect to rental price of skill of roughly 5.6. However, this is somewhat of an exaggeration, as some of the increase in full-time work must come from reduced part-time work. Unfortunately, Francesconi (2002) does not report the resulting decrease in part-time work that accompanies.

The last two papers on female labor supply described below are Keane and Wolpin (2010). In these papers, Keane and Wolpin utilize approximate solution methods developed in Keane and Wolpin (1994), and estimation methods developed in Keane and Wolpin (2001), to estimate a model of female life-cycle behavior that is considerably richer than previous models in the literature. Both marriage and fertility are treated as choices, and both full and part-time work options are available. Schooling is also a choice. An important feature of the data that is not accommodated in prior dynamic models is that a large fraction of single women with children participate in public welfare programs. Thus, welfare participation (when eligible) is also incorporated as a choice.

In the model, women begin making decisions at age 14, and the terminal period is age 65. The fertile period is assumed to last up until age 45, and during this period women have up to 36 choice options in each period. Afterwards they have up to 18 options.[91] The

---

[90]Neither model captures the sharp decline in participation in the 60s due to retirement. But to be fair neither model incorporates any features designed to explain retirement behavior (such as pensions or Social Security).

[91]The choice set differs across women for a number of reasons. For instance, only unmarried women with children under 18 have the option to participate in welfare, and working while on welfare is not an option if

decision period is assumed to be 6-months until age 45, which is a compromise between the length of a school semester and the child gestation period. After age 45, the decision period is one year (as the fraction of women who either attend school or have children after 45 is negligible). Given that behavior of girls as young as 14 is being modelled, it is essential to consider the role of parental co-residence and parental income support. Yet, as this is not a focal point of the model, the authors choose not to treat living with parents as a choice. Both the probability of co-residence and parental transfers are treated as stochastic processes that depend on a person's state variables.

One fundamental difference from van der Klaauw (1996) and Francesconi (2002) is that marriage is treated as a true search process. Each period a woman may receive a marriage offer that consists of: (1) the mean wage of the husband, and (2) a marriage quality draw (which captures non-pecuniary aspects of the match). The potential husband's mean wage depends on the woman's characteristics, such as her schooling and skill level, as well as a permanent component drawn from a distribution. Thus, a husband fixed effect becomes part of the state space. In this setup, a woman has an incentive to reject marriage offers while waiting for a husband with a high mean wage.

Another fundamental difference from prior work is that the model is non-stationary in the sense that the economic environment changes over time. Specifically, the welfare rules change over time and differ by state, so each cohort of women (as defined by the semi-annual period in which they reach age 14) in each state faces a different sequence of welfare rules. This creates a number of computational problems. First, each cohort of women in each state faces a different dynamic optimization problem (raising computational burden). Second, one must make an assumption about how women forecast future rules. Third, the rules are complex, making it difficult to characterize them.

Keane and Wolpin (2010) deal with these problems as follows. First, they develop a simple 5 parameter function that characterizes the welfare benefit rules in each State in each year quite accurately. Second, they assume women use a State-specific VAR in these 5 parameters to predict future rules. Third, they only use data from 5 large States, so as to

---

the offer wage rate is high enough that income would exceed the eligibility level. Also, girls under 16 cannot choose marriage.

reduce the number of DP problems that must be solved in estimation. This enables them to use the data from other states for out-of-sample validation.

Keane and Wolpin (2007, forthcoming) assume that a woman receives disutility from a variable that measures "non leisure" time. This is a sum of work hours, a fixed time cost of work, time spent in school, time required to collect welfare, and time required to care for children.[92] The authors estimate weights on the variables other than work hours to account for the fact that school time, child care time and time collecting welfare may entail more/less disutility than time spent working. A woman receives utility from consumption, which is assumed to be a share of total household income. Utility is quadratic in non-leisure time and linear in consumption. Similar to the previous papers we discussed, consumption is interacted with non-leisure time. The estimated coefficient is negative, implying that consumption and leisure are complements, inducing negative income effects on labor supply and fertility.

Additional interactions are introduced that allow marriage and children to shift the degree of complementarity between consumption and leisure. This would have been irrelevant in the papers discussed previously, as they do not try to explain labor supply, marriage and fertility choices jointly. The estimates imply that marriage and children both significantly reduce the degree of complementarity between consumption and leisure, but do not eliminate it.

Women also receive utility/disutility from children, pregnancy, marriage, school attendance and welfare participation. Utility is quadratic in number of children. The utility/disutility from pregnancy is a polynomial in age. As one would expect, this becomes a large negative for women as they approach 45, consistent with the greater risks associated with pregnancy at older ages. The disutility of welfare attendance enables the model to explain the common phenomenon of non-participation by eligible women (see Moffitt (1983)). The utility function coefficient on each of the 5 choice variables (hours, pregnancy, marriage, school and welfare) consists of a constant plus a stochastic taste shock. This enables the model to generate a non-zero probability of any observed choice outcome.

The model allows for unobserved heterogeneity in the form of 6 types of women who differ

---

[92]Childcare time is, in turn, a weighted sum of time required to care for children in different age ranges.

in the preference parameters (constant terms) associated with the 5 choice variables (i.e., different tastes), and in the intercepts of the own and potential husband offer wage functions (i.e., different skills). The model includes observed heterogeneity as well; the heterogeneous skill and taste parameters differ across states and across ethnic groups (blacks, whites and Hispanics). Finally, the utility function includes interactions of indicators for full and part-time work, school and marriage with lagged values of these indicators, to capture state dependence in tastes for these choice options.[93]

The model is estimated using data from the National Longitudinal Survey of Youth 1979 cohort (NLSY79). The NLSY79 includes women aged 14 to 21 in 1979. The paper uses the data from the years 1979 to 1991. Thus, the women reach a maximum age of 33. The states used in estimation are California, Michigan, New York, North Carolina and Ohio. To be in the sample, a woman had to reside in the same state for the whole sample period, which screens out about 30%. This leaves data on approximately 2800 women.[94] The annual discount factor is fixed at 0.93.

Estimates of the log wage function imply that (at the mean of the data) an additional year of school raises wages by 9.1%. And 84% of the variance of wages is attributed to measurement error (the true log wage standard deviation is 0.17). The experience coefficients imply that the first year of full-time work raises wages by 2.6%, and that the experience profile peaks at 36 years. In addition, lagged full-time work raises the current wage offer by 7%, while lagged part-time raises it by 3%. Black and Hispanic women have lower offer wages than white women (by 13% and 6%, respectively).

In the husband offer wage function, the coefficient of the woman's skill endowment (i.e.,

---

[93]The utility function includes some miscellaneous additional terms that were added to capture some specific features of the data. Full and part-time work are interacted with school to capture the fact that people who work while in school tend to have a strong preference for part-time over full-time work. Work variables are also interacted with a school less than 12 dummy to capture that part-time work is far more prevalent among high school students. Pregnancy is interacted with school to capture that women rarely go to school while pregnant. Tastes for school, marriage and pregnancy are also allowed to shift at certain key ages (16, 18 and 21). And there is a linear time trend (across cohorts) in tastes for marriage.

[94]Keane and Wolpin (2002), which presents a nonstructural analysis of the same data, provides a more detailed description.

intercept in the woman's wage function) is 1.95, implying a very high degree of assortative mating on skill. And each additional year of education for the woman raises the husband offer wage by 3%., Black and Hispanic women have much lower husband offer wages than whites (by 30% and 14%, respectively). The estimates imply that women receive 55% of total household income. So, just as in van der Klaauw (1996) and Francesconi (2002), much of the return to schooling appears to emerge through the marriage market.

Keane and Wolpin (2007) provide a good deal of evidence on the fit of the model and assess how well it predicts behavior in the holdout state of Texas. The model performs reasonably well in these tests, including providing better predictions than some candidate competing nonstructural models.

As has been the focus of the labor supply literature, Keane and Wolpin (2009) estimate labor supply wage elasticities. Recall that the model has six types of women, which we can rank by skill level from type 1s (highest skill endowment) to types 6s (lowest). Type 6s account for the majority of welfare participants. Keane and Wolpin (2010) report experiments where they increase the offer wage by 5% for each type separately. The wage elasticities are inversely proportional to skill level, ranging from only 0.6 for type 1s to 9.2 for type 6s. Thus, the overall elasticity of 2.8 is deceptive with regard to behavior of various subsets of the population.

For type 6 women, the 5% wage increase has a dramatic impact on all aspects of their behavior. For instance, for white women of type 6, the percent working at ages 22 to 29.5 increases from 34% to 50% (a 47% increase). But it is also notable that mean completed schooling increases from 11.5 to 12 years, the high school drop out rate drops from 42% to 24%, welfare participation drops from 25% to 20%, and incidence of out-of-wedlock teenage pregnancies drops from 3.4% to 2.8%. All of these behavioral changes (i.e., more education, fewer teenage pregnancies, less welfare participation) contribute to the increase in labor supply. In contrast, type 1s are already completing a high level of schooling, are rarely having children at young ages, are not participating in welfare, and are participating in the labor market at a high rate. Thus, in a sense there are fewer channels through which a wage increase can affect them. In summary, the results indicate that wage elasticities of labor supply for low skilled women are much greater than for high skilled women.

It is difficult to summarize the estimates of labor supply elasticities for women across the studies we have surveyed. Several of the non_DCDP studies we have examined calculate what might be called "short run" elasticities that hold work experience, marriage and fertility fixed. On the other hand, the DCDP models calculate "long run" elasticities that allow, depending on the study, some combination of experience, fertility, marriage and education to adjust to wage changes. Nevertheless, a reasonable assessment of the estimates from this literature is that the labor supply elasticity estimates for women are generally quite large. The DCDP models give uniformly large "long run" elasticities ranging from 2.8 to 5.6. The life-cycle models of Heckman and MaCurdy (1982) and Kimmel and Kniesner (1998) give large Frisch elasticities (2.35 to 3.05). The Marshallian elasticity of 0.89 obtained by Cogan (1981) in a static model is also quite large.[95] Thus, 7 of the 9 studies obtain large female labor supply elsaticities (of various types). Only the Blundell and Walker (1986) and Blundell, Duncan and Meghir (1998) studies find small elasticities. This may be because these two studies consider the labor response of working women to wage changes, while the other 7 studies incorporate the participation margin.

The richness of the Keane and Wolpin (2010) model enables them to address a variety of substantive issues beyond calculating labor supply elasticities. These focus on (i) the factors that account for differences between blacks, whites and Hispanics in choice behavior and (ii) the effects of changing welfare rules. With respect to behavioral differences among minority and white women, the model estimates indicate that black women face a worse marriage market than do white women. The mean earnings of potential husbands, that is, the pool of men who make marriage offers, is 27 percent lower for black than for white women. In addition, unobservable traits of potential mates reduce the psychic value of getting married by $2,500 (in 1987 NY dollars) for black women relative to white women. The estimates also indicate that black women face poorer labor market opportunities. Wage offers are 12.5 percent lower for black women than for white women. In terms of preferences, the stigma attached to being on welfare is smaller for black women, although the difference, 290 dollars per six month period, does not seem that large. Black women do not differ from white women

---

[95]Note that this is an elasticity for hours conditional on working. It is unfortunate that Cogan does not report a participation elasticity, as, given his estimates, this would presumably have been much larger.

in the disutility they attach to work (an extra 1000 hours of work is equivalent to a 117 dollar greater drop in consumption for black than for white women), but they are estimated to have a significantly greater preference for children (the birth of a child is equivalent to a greater increase in consumption by 1,352 dollars for black women than for white women).

To assess the importance of labor market, marriage market and preference differences, Keane and Wolpin (2009) simulate behaviors of black women under alternative counterfactual scenarios. They find that equalizing marriage market opportunities between black and white women would reduce welfare participation of black women at, for example, ages 26-29 from 29.7 percent to 21.4 percent, thus closing 37 percent of the black-white gap. Equalizing labor market opportunities has a somewhat larger impact, reducing the gap by about 45 percent. However, these changes have opposite effects on employment. Improving marriage market opportunities of black women in this age group reduces their employment rate from 55.7 to 42.3 percent, and thus widens the black-white gap, while employment rates are essentially equalized when labor market opportunities are equalized. Both counterfactuals increase marriage rates in that age range, although directly operating on marriage market conditions has a much larger impact, reducing the marriage rate gap of 37 percentage points to only 10 percentage points. Along with this large increase in marriage rates, the mean number of teenage births increase slightly. On the other hand, the relatively small increase in marriage rates that accompany the counterfactual improvement in the labor market leads to a fall in the mean number of teenage births by 13 percent and closes the black-white gap by 38 percent. Finally, improving the marriage market of black women reduces their completed schooling by a third of a year on average, widening the gap with white women, while improving their labor market opportunities increases their completed schooling by 0.2 years. Increasing welfare stigma of black women to that of white women, given the relatively small difference noted above, has only a small impact on behavior; the largest effect is to reduce welfare participation by 3.3 percentage points, significantly less than that exhibited for the other counterfactuals.

As these counterfactual experiments illustrate, none of these differences, when taken one at a time, can account for the racial differences in outcomes. Improving marriage market opportunities, by itself, reduces some of the gaps, but widens others. Improving labor market

opportunities reduces all of the gaps, but considerably less so for demographic outcomes. And, welfare stigma accounts for little of the racial differences in behavior.

In another counterfactual experiment, Keane and Wolpin (2010) simulate the effect of eliminating welfare. Because welfare receipt is heavily concentrated among one of the six (unobserved) types, this experiment was performed only for women with the preferences and opportunities of the women of that type. For this type, 68.1 percent of black women and 24.6 percent of white women are receiving welfare at ages 26-29. That difference, of 43.5 percentage points is eliminated in the experiment. But, perhaps the most striking result is that eliminating welfare also essentially eliminates the employment gap, even though labor market opportunities are worse for black women. The original gap of 16.5 percentage points is reduced to 1.4 percentage points. Eliminating welfare also increases marriage rates more for black women, by 14.8 percentage points, than for white women, 8.2 percentage points, reducing the original gap from 36.2 percentage points to 29.6 percentage points. The mean number of teenage births fall slightly, but about the same for black and white women. A similar result is observed for the proportion of women of this type who do not graduate from high school.

As Keane and Wolpin (forthcoming) conclude, there is no simple answer as to what causes the differences in the behavior of black and white women. The welfare system in place in the U.S. until the major reform in 1996 differentially affected the labor market attachment of black women, but did not by itself account for much of the difference in marriage, fertility and schooling. The poorer marriage and labor market opportunities of black women both contributed importantly to the greater dependency of black women on welfare. Ultimately, it is the interaction of all of these factors, the welfare system, opportunities and preferences that jointly account for the large racial gaps in labor market and demographic outcomes.

In summary, the female labor supply literature has emphasized the connection between participation decisions and human capital, fertility and marriage. Those papers that have attempted to model fertility and/or marriage as choices have ignored savings behavior to achieve computational tractability. There is as of yet no model of female life cycle behavior that includes savings along with human capital, fertility and marriage. This is an important avenue for future research, although a difficult one, because it involves modeling interactions

within a household in a dynamic framework.[96]

5.   Male Labor Supply

As we have noted, DCDP models of female labor supply have ignored considerations of consumption smoothing through savings and borrowing behavior. This is in sharp contrast to the literature on male labor supply, which has made consumption smoothing a major focus, in conjunction with a continuous hours choice and, with few exceptions, has ignored human capital accumulation.

Indeed, without availing itself of the DCDP approach, the literature on males has generally adopted estimation methodologies that specifically seek to avoid having to solve the full dynamic programming problem. A notable example is the seminal work by MaCurdy (1981, 1983), who developed estimation methods using the Euler conditions of dynamic models of labor supply with savings. Shaw (1989) extended this approach to a model with labor supply, savings and human capital accumulation. Of course, the DCDP methodology does not preclude modeling all these aspects of behavior, but it is computationally burdensome. On the other hand, a limitation of Euler equation and other "non-DCDP" or "non-full solution" approaches, is that, while they can deliver structural parameter estimates, they do not in general allow one to simulate behavioral responses to changes in policy (or the economic environment more generally).

To our knowledge only one paper, Imai and Keane (2004), has used DCDP methodology to estimate a labor supply model with assets, continuous hours and on-the-job human capital accumulation. We present a simplified version of the Imai and Keane (2004) model that captures the main points.

Assume that a worker's human capital, denoted by K, evolves according to the simple human capital production function

$$K_{i,t+1} = (1 + \alpha h_{it})K_{it}. \tag{120}$$

The growth in human capital, in this formulation, is a constant fraction of hours worked.

_____

[96]For recent attempts see Mazzocco and Yamaguchi (2007) and Tartari (2006).

K_{i1} is the person's skill (or human capital) endowment at the time of labor force entry.[97]
A person's wage at time t, w_{it}, is equal to the current stock of human capital times the
(constant) rental price of human capital, R.[98] Human capital is subject to a transitory
productivity shock. Specifically,

$$w_{it} = RK_{it}(1 + \epsilon_{it}). \tag{121}$$

The period-specific utility function is given by

$$U_t = \frac{C_t^{1+\eta}}{1+\eta} + \beta_t \frac{h_t^{1+\gamma}}{1+\gamma}, \quad \eta \le 0, \ \ \gamma \ge 0, \tag{122}$$

where β_t is an age-varying parameter that shifts tastes for work. In contrast to the female
labor supply literature, where utility is typically assumed to be linear in consumption, utility
in (122) is CRRA in consumption. Given the emphasis of the male literature on savings, the
CRRA is a more natural choice. Assets evolve according to

$$A_{t+1} = (1+r)(A_t + (1-\tau)w_t h_t - C_t) \tag{123}$$

where τ is the tax rate on labor income. Given this setup, the state space at t, Ω_t, consists
of $\{K_t, A_t, \epsilon_t, \beta_t\}$.

The individual is assumed to maximize the expected present discounted value of utility
over a finite horizon. The value function at age t is then

$$
\begin{aligned}
V_t(\Omega_t) &= \max_{C_t, h_t} U_t + E_t \left(\sum_{j=t+1}^{T} \delta^{j-t} U_j \right) & (124) \\
&= \max_{C_t, h_t} U_t + \delta E(V_{t+1}(\Omega_{t+1})|\Omega_t, C_t, h_t), & (125)
\end{aligned}
$$

where the expectation is taken over the future transitory productivity shocks and tastes for
work conditional on the current state space. As is common in these types of models, we
assume these stochastic terms are independent over time. In that case, we can replace Ω_t

---

[97]Imai and Keane (2004) actually assume a much more complex process, designed to capture patterns of
complementarity between human capital and hours of work in the human capital production function. But
use of this simpler form helps to clarify the key points.

[98]Although Imai and Keane assume a constant rental price, allowing for time varying rental rates is fairly
straightforward.

with Ω_t^-, that is, we can drop ϵ_{it} and β_t from Ω_t in forming the expectation in (125). Note that $E(V_{t+1}(\Omega_{t+1}^-, \epsilon_{t+1}, \beta_{t+1})|\Omega_t^-, C_t, h_t)$ is simply the analog of the $E\max_{t+1}$ function in the discrete choice problem already discussed.

As in the discrete choice case, the solution of the model consists of finding the $E\max_t$ functions. Imai and Keane (2004) do that using a backsolving and approximation procedure similar to Keane and Wolpin (1994), adapted to continuous choice variables. In the terminal period, the value function is

$$V_T(\Omega_T) = \max_{C_T, h_T} \left(\frac{C_T^{1+\eta}}{1+\eta} + \beta_T \frac{h_T^{1+\gamma}}{1+\gamma} \right). \tag{126}$$

In this simple static problem, and without a bequest motive, given w_T and A_T, the consumer chooses C_T and h_T to maximize utility subject to the budget constraint $C_T = (1-\tau)w_T h_T + A_T$.[99]

In principle, the backsolving procedure starts by calculating $V_T(\Omega_T)$ for every possible state in Ω_T at which the worker might enter period T. The solution for h_T is given by the first-order condition

$$\frac{\beta_T h_T^\gamma}{[(1-\tau)w_T h_T + A_T]^\eta} = (1-\tau)w_T \tag{127}$$

This equation can be solved numerically for the optimal h_T using an iterative search procedure. Once the optimal h_T is determined for each state point, the optimal C_T is found from the budget constraint. $V_T(\Omega_T)$ is then found by substituting the optimal value of h_T and C_T into (126).

Although we need to calculate $E\max_T$ only at the deterministic components of the state space, Ω_T^-, a problem arises, because the number of possible levels of human capital and assets at the start of period T is extremely large, if not infinite. Thus, it is not computationally feasible to literally solve for $E\max_T$ for every possible state value. Thus, Imai and Keane (2004) adopt the Keane and Wolpin (1994) approximation method discussed earlier, which involves solving for $E\max_T$ at a finite (and relatively small) subset of the possible state points.

To implement that procedure, a regression is estimated as some flexible function of the state variables and used to predict or interpolate the value of $E\max_T$ at any desired state

---

[99] Adding a bequest motive to the model, as in Imai and Keane (2004) is straightforward. This extension can be accommodated by adding a terminal value function, say $f(A_{T+1})$ to $V_T(\Omega_T)$..

point (K_T, A_T), including, in particular, points that were not among those used to fit the regression. Thus, having fit this interpolating regression, we may proceed as if $E \max_T$ is known for every possible state point in (K_T, A_T). As before, denote the interpolating function that approximates $E \max_T$ as $\widehat{E \max}_T$. We must assume that $\widehat{E \max}_T$ is a smooth differentiable function of K_T and A_T (e.g., a polynomial) for the next step. For expositional convenience, let $\widehat{E \max}_T$ be the following simple function,

$$\widehat{E \max}_T = \pi_{T0} + \pi_{T1} \log K_T + \pi_{T2} \log A_T + \varsigma_T, \tag{128}$$

and let $\pi_T(K_T, A_T) = \widehat{\pi}_{T0} + \widehat{\pi}_{T1} \log K_T + \widehat{\pi}_{T2} \log A_T$ be the predicted value of $\widehat{E \max}_T$, where the $\widehat{\pi}'s$ are estimated parameters.

As in the discrete choice setting, the next step of the backsolving process moves back to period $T - 1$. Then, using the predicted values of $E \max_T$ from the approximating function,

$$V_{T-1}(\Omega_{t-1}) = \max_{C_{T-1}, h_{T-1}} U_{T-1} + \delta \pi_T(K_T, A_T). \tag{129}$$

Upon substituting in the laws of motion for K_T and A_T, we get

$$
\begin{aligned}
V_{T-1}(\Omega_{t-1}) = \max_{C_{T-1}, h_{T-1}} & \frac{C_{T-1}^{1+\eta}}{1+\eta} + \beta_{T-1} \frac{h_{T-1}^{1+\gamma}}{1+\gamma} \\
& + \delta(\widehat{\pi}_{T0} + \widehat{\pi}_{T1} \log((1 + \alpha h_{i,T-1}) K_{i,T-1}) \\
& + \widehat{\pi}_{T2} \log((1 + r)(A_{T-1} + (1 - \tau) w_{T-1} h_{T-1} - C_{T-1}).
\end{aligned}
\tag{130}
$$

Finding the optimal values of consumption and hours is now just like a static optimization problem. The first order conditions are given by

$$
\begin{aligned}
\frac{\partial V_{T-1}}{\partial C_{T-1}} &= C_{T-1}^{\eta} - \frac{\delta \widehat{\pi}_{T2}}{A_{T-1} + (1 - \tau) w_{T-1} h_{T-1} - C_{T-1}} = 0, \\
\frac{\partial V_{T-1}}{\partial h_{T-1}} &= -\beta_{T-1} h_{T-1}^{\gamma} + \frac{\delta \alpha \widehat{\pi}_{T1}}{1 + \alpha h_{T-1}} + \frac{\delta(1 - \tau) w_{T-1} \widehat{\pi}_{T2}}{A_{T-1} + (1 - \tau) w_{T-1} h_{T-1} - C_{T-1}} = 0.
\end{aligned}
\tag{131}
$$

These two equations can be solved numerically for C_{T-1} and h_{T-1} at any given state point in Ω_{t-1}.[100]

Following the development for period T, the next step is to calculate the values of $E \max_{T-1}$ at a subset of the state points. For a given value of K_{T-1} and A_{T-1}, we can

---

[100]These equations may have multiple solutions If there are, then one would need to check second order conditions or calculate V_{T-1}.

substitute the optimal C_{T-1} and h_{T-1} into V_{T-1} ((130)) and numerically integrate over the joint distribution of ϵ and β. Given the values of $E\max_{T-1}$, we can then estimate the interpolating function at $T-1$, say

$$\widehat{E\max}_{T-1} = \pi_{T-1,0} + \pi_{T-1,1}\log K_{T-1} + \pi_{T-1,2}\log A_{T-1} + \varsigma_{T-1}, \tag{132}$$

Using this interpolating function, we can write the (approximate) value functions at time $T-2$ in analogous fashion to ((130)). The only difference is in the interpolating function parameters. These steps are repeated until an approximate solution is obtained for every period back to $t = 1$.

The approximate solution consists of the complete set of interpolating function parameters, the $\widehat{\pi}_t$'s for $t = 2, ..., T$. Given these estimated interpolating functions, it is possible to solve numerically the simple two equation system like (131) at each t to find the optimal choice of a worker at any point in the state space. In particular, using $\widehat{E\max}_2$, one can solve for optimal labor supply and consumption in period $t = 1$, the first period of the working life. As previously noted, this is what first order conditions alone do not provide. Furthermore, by drawing values for the taste shocks and rental rates and repeatedly solving optimal labor supply and consumption over time, one can simulate entire career paths of workers. This enables one to simulate how changes in the economic environment, such as changes in tax rates, would affect the entire life-cycle path of labour supply and consumption, as one can re-solve the model and simulate career paths under different settings for the policy parameters.

Imai and Keane (2004) estimate their model using white males from the NLSY79. They choose this data set because of its fairly extensive asset data. The men in their sample are aged 20 to 36 and, as the focus of their paper is solely on labor supply, they are required to have finished school. Due to the computational burden of estimation they randomly choose 1000 men from the NLSY79 sample to use in estimation. People are observed for an average of 7.5 years each starting from the age at school completion.

Notably, Imai and Keane (2004) allow for measurement error in observed hours, earnings and assets when constructing the likelihood function. As all outcomes are measured with error, construction of the likelihood is fairly simple. One can simulate career histories for

each worker, and then form the likelihood of a worker's observed history of hours, earnings and assets as the joint density of the set of measurement errors necessary to reconcile the observed history with the simulated data.[101]

Imai and Keane (2004) estimate that $\gamma = 0.26$. In a model without human capital, this would yield a Frisch elasticity of $\frac{1}{\gamma} = 3.8$, which implies a much higher willingness to substitute labor intertemporally than in almost all prior studies for men (see MaCurdy (1983) for an exception). Simulations of the model reveal that, even accounting for human capital effects, the estimate of γ implies more elastic labor supply than in most prior work.

Imai and Keane (2004) explain their high estimate of intertemporal substitution based on the logic of Figure 1. The figure presents a stylized (but fairly accurate) picture of how wages and hours move over the life cycle. Both wages and hours have a hump shape, but the hump in wages is much more pronounced. This apparently weak response of hours to wages leads conventional methods of estimating the intertemporal elasticity of substitution (which ignore the effect of working on the accumulation of human capital) to produce small values.

Indeed, Imai and Keane (2004) show that if they simulate data from their model and apply instrumental variable methods like those in MaCurdy (1981) and Altonji (1986) to estimate $\frac{1}{\gamma}$, they obtain values of .325 (standard error = .256) and .476 (standard error = .182), respectively. Thus, the Imai-Keane (2004) model generates life-cycle histories that, when viewed through the lens of models that ignore human capital accumulation, imply similarly low co-movement between hours and wages to those obtained in most prior work. As further confirmation of this point, the authors report simple OLS regressions of hours changes on wage changes for both the NLSY79 data and the data simulated from their model. The estimates are -0.231 and -0.293, respectively. Thus, a negative correlation between hours changes and wage changes in the raw data is perfectly consistent with a high willingness to substitute labor intertemporally over the life cycle.

What reconciles these prima facie contradictory observations is the divergence between the opportunity cost of time and the wage in a model with returns to work experience. In particular, Imai and Keane (2004) estimate that from age 20 to 36 the mean of the

---

[101]Keane and Wolpin (2001) first developed this appraoch to forming the likelihood in DCDP models. Keane and Sauer (2009) extended the approach to nonstructural panel data models.

opportunity cost of time increases by only 13%. In contrast, the mean wage rate increases by 90% in the actual data, and 86% in the simulated data. Thus, the wage increases about 6.5 times faster than the opportunity cost of time. These figures imply that conventional methods of calculating $\frac{1}{\gamma}$ will understate it by a factor of roughly 6.5.

This point is illustrated in Figure 1 by the line labelled "wage + hc," which adds the wage and the return to an hour of work experience (in the form of higher future earnings) to obtain the opportunity cost of time. As the Figure illustrates, the opportunity cost of time is much flatter over the life cycle than is the wage rate. Thus, hours appear to be much more responsive to changes in the opportunity cost of time than to changes in wages alone.

Imai and Keane (2004) use their model to simulate how workers of different ages would respond to a 2% temporary unanticipated annual wage increase. For a worker at age 20, hours increase only 0.6%. But the response grows steadily with age. At age 60 the increase in hours is nearly 4%, and at age 65 it is about 5.5%. The reason the effect of a temporary wage increase rises with age is that, as depicted in Figure 1, as a person ages the current wage becomes a larger fraction of the opportunity cost of time. According to Imai and Keane (2004)'s estimates, at age 20 the wage is less than half of the opportunity cost of time, but by age 40 the wage is 84% of the opportunity cost of time.

Unfortunately, the Imai and Keane (2004) simulations do not reveal what the model implies about how workers would respond to permanent tax changes. To fill this gap, Keane (2009a) uses the Imai-Keane model to simulate the impact of a permanent 10% tax rate increase (starting at age 20 and lasting through age 65) on labor supply over the entire working life. If the tax revenue is simply thrown away, the model implies that average hours of work (from ages 20 to 65) drops from 1992 per year to 1954 per year, a 2% drop. If the revenue is redistributed as a lump sum transfer, labor supply drops to 1861 hours per year, a 6.6% drop. The later is as a reasonable approximation to the compensated elasticity with respect to permanent tax changes implied by the model (i.e., 0.66).

The effects of the tax, however, are very different at different ages. As seen in Table 2, tax effects on labor supply slowly rise from age 20 to about age 40. Starting in the 40s, the effects on labor supply start to grow quite quickly, and by age 60 effects are substantial. Thus, in response to a permanent tax increase, workers not only reduce labour supply, but

also shift their lifetime labor supply out of older ages towards younger ages.

To our knowledge, there are only two papers besides Imai and Keane (2004) that have used full solution methods to estimate a life-cycle model that includes both human capital investment and savings, Keane and Wolpin (2001) and van der Klaauw and Wolpin (2008). Neither of those papers, however, models the continuous choice of hours, although they allow for several discrete alternatives.[102] The main focus of the Keane and Wolpin (2001) paper is on schooling choice (not labor supply), so we discuss it in a later section. But their paper is of interest here because it assumes a CRRA utility function in consumption, and so, like Imai-Keane, provides an estimate of the key preference parameter η, which governs income effects in labor supply and inter-temporal substitution in consumption. Keane and Wolpin (2001) obtain $\eta \approx -0.50$, which implies weaker income effects, and less curvature in consumption (i.e., higher willingness to substitute inter-temporally), than much of the prior literature. Keane and Wolpin (2001, p. 1078) argue that the reason is that their work accommodates liquidity constraints, and that failure to do so may have led to a downward bias in estimates of η in prior work.[103]

Imai and Keane (2004) estimate that $\eta = -0.74$. This implies a somewhat lower intertemporal elasticity of substitution in consumption than the Keane and Wolpin (2001) estimate of $\eta \approx -0.50$(that is,$\frac{1}{\eta} = -1.35$ vs. -2.0). But their estimate of η still implies weaker income effects on labor supply, and a higher willingness to substitute consumption inter-temporally, than much of the prior literature. Instead of liquidity constraints (as in Keane and Wolpin (2001)) the Imai and Keane (2004) model "explains" the fact that young workers do not borrow heavily against higher future earnings by assuming age effects in the marginal utility of consumption. Both models provide a good fit to asset data over the life-cycle. Finally, Keane (2009b) uses the Imai-Keane estimates of γ and η to calibrate a simple two-period equilibrium model. He finds that welfare costs of labor income taxation

---

[102]Van der klaauw and Wolpin (2008) estimate a collective model of the joint labor supply decisions of a married couple nearing retirement They allow for savings and human capital accumulation, incorporating as well a detailed representation of U.S. social security system.rules. As noted in the introduction, we do not review the DCDP retiremnent literature in this chapter.

[103]The estimate of η in van der Klaauw and Wolpin (2008) is -.6, which is also in line with other estimates from the retirement literature. They also include liquidity constraints.

are much larger than more conventional values of γ and η would suggest.

In summary, although the literature that uses dynamic programming models to study life-cycle labor supply, asset accumulation and human capital investment for males is quite small, it has produced important results. Specifically, it finds that the intertemporal elasticities of substitution for both labor supply and consumption are quite a bit larger than implied by earlier work. This, in turn, implies that tax effects on labor supply for males may be larger than conventionally thought. Clearly more work is called for to investigate the robustness of these results to alternative model specifications and data sources.

B.   Job Search

Along with the dynamic labor force participation model, among the first applications of the DCDP approach was to the estimation of models of job search - the transition from unemployment to employment. The labor supply and job search literatures have, however, addressed different questions and followed distinct paths. To understand why that has been the case, recall that in the labor force participation model workers with the same characteristics, and thus the same level of productivity, are offered the same wage. That is, in the wage offer function, $\log w_{it} = \log r + \gamma_0 + z_{it}\gamma + \eta_{it}$, r is assumed to be a market-level (for example, competitively determined) skill rental price and z_{it} and η_{it} are worker characteristics.

In contrast, the job search literature starts from the assumption that firms may offer different wages (skill rental prices, r) to identical workers within a given labor market. Then, the wage offer received by a worker of given characteristics from a firm j is $\log w_{itj} = \overline{\log r} + \gamma_0 + z_{it}\gamma + \zeta_j$, where $\overline{\log r}$ is the mean skill rental price in the labor market and where ζ_j reflects firm $j's$ idiosyncratic component of the skill rental price. In the basic model, the accepted job lasts "forever" and individuals are not subject to productivity shocks. Given this wage structure, once an individual accepts a wage offer from a firm with a given ζ_j, their skill rental price is fixed for as long as they work for that firm.[104] The information set of the individual includes the distribution of ζ_j across firms, but not which firms are matched with particular values of ζ. Because there are more and less desirable firms, individuals

---

[104] Wages would grow deterministically if z contains age or job tenure.

have an incentive to engage in job search, that is, to look for a high wage firm. Job search is sequential. The difference in the labor force participation and job search models thus reflects the different assumptions made about the wage structure of the labor market.

The partial equilibrium search model has normally been used to understand a different phenomenon than has the labor supply model. Official labor force statistics distinguish among three mutually exclusive and exhaustive states, being employed, being unemployed and being out-of-the labor force. The distinction between the latter two states is based on whether an individual is actively seeking work. Both the labor force participation and search models consider only two states. In the labor force participation model, unemployment and out-of-the labor force are collapsed into one nonemployment state. In that model, it is assumed that a new wage offer is received every period and any individual will work at some offered wage. The search model conditions on individuals having already chosen unemployment over being out-of-the labor force and does not assume that a job offer necessarily arises each period. Because of this difference, labor force participation models have been applied to low frequency data based on the employment-nonemployment dichotomy, commonly at the annual level and often for women, while job search models have been applied to high frequency data, for example, at the weekly level, based on the employment-unemployment dichotomy.

The structural implementation of the standard partial equilibrium job search model was first considered by Wolpin (1987) and van den Berg (1990), building upon a non-structural literature that had begun a decade or more before.[105] The non-structural empirical literature was focused on the evaluation of the effect of UI programs, more specifically, on the estimation of the impact of unemployment benefits on the duration of unemployment and wages. The empirical approach in that literature was (and is still) based, loosely and in some ways incorrectly, on the sequential job search model first formalized by McCall (1970) and Mortensen (1970). The structural empirical literature, following the DCDP paradigm, is based on the explicit solution and estimation of the sequential model.

---

[105]A number of these earlier papers appeared in a 1977 symposium volume of the *Industrial and Labor Relations Review*. Most relevant in that volume are papers by Classen and Holen, and the comment on them by Welch.

There have been many extensions and modifications of the standard model in the structural empirical literature. Within the standard framework, Stern (1989), extending the original contribution of Stigler (1961) to a sequential framework, allowed for simultaneous search, that is, for the submission of multiple job applications in a period. Blau (1991) dropped the assumption of wealth maximization, allowing for job offers to include not only a wage offer but also an hours offer. Ferrall (1997) incorporated all major features of the Canadian UI program. Gemici (2007) considered the joint husband and wife search-migration decision in an intra-household bargaining framework. Paserman (2008), adopting a behavioral approach, allowed for hyperbolic discounting.

The standard model has also been extended beyond the consideration of the single transition from unemployment to employment. Wolpin (1992) incorporated job-to-job transitions and both involuntary and voluntary transitions into unemployment.[106] Rendon (2006) allowed for a savings decision in a setting where agents can also transit, both through quits and layoffs, from employment to unemployment.[107] Both of these latter papers also allowed for wage growth with the accumulation of work experience, employer-specific (tenure) in the case of Rendon and both general and employer-specific in the case of Wolpin.

As noted, the standard job search model assumes that *ex ante* identical workers may receive different wage offers, or analogously, that the same unemployed worker may receive different offers over time. Diamond (1970) showed that with the assumptions of the standard job search model, in a game in which firms are aware of worker search strategies, the wage offer distribution will be degenerate at the worker's reservation wage or outside option. This result led to the development of models in which wage dispersion could be rationalized as an equilibrium outcome, which, in turn led to a structural empirical literature. The empirical literature has focused on two kinds of models, those based on a search-matching-bargaining

---

[106]Burdett (1978) extended the standard unemployment search model to allow for search on the job.

[107]It is a common theme in the structural literature to build upon and extend the theoretical literrature in developing estimable models. This is the case with Rendon's (2006) paper, which builds on the earlier work of Danforth (1979). Lentz (2009) also structurally estimates a sequential search model with savings. Unlike the standard model, the wage offer distribution is taken to be degenerate and agents choose their search intensity, which affects the rate at which job offers are received. Lentz uses the model to empirically detemine the optimal unemployment insurance scheme.

approach to wage determination (Diamond and Maskin (1979), Mortensen (1982), Wolinsky (1987)) and those based on wage-posting by firms that gain monopsony power through search frictions (Albrecht and Axell (1984) and Burdett and Mortensen (1998)). These models, not only rationalized wage dispersion, but also allowed for quantification and policy analyses in an equilibrium setting, for example, changes in UI benefits or changes in the level of the minimum wage.

The empirical implementation of equilibrium search models has become a major strand of the structural job search literature.[108] Embedded within those models are different variants of the standard partial equilibrium search model and, in that sense, the development of the DCDP estimation approach was a critical precursor. However, given that the modeling has gone well beyond the partial equilibrium search model to which the DCDP approach has direct application, it would take us too far afield to provide a review of that literature. For such a review, we would refer the reader to the chapter by Mortensen and Pissarides in the Handbook of Labor Economics (Volume 3b, 1999) or the more recent survey by Eckstein and Vandenberg (2007).

In the rest of this section we review the structural empirical literature on the partial equilibrium job search model. Because the structural literature is explicitly connected to the theory, we first present the formal structure of the standard job search model and show how the nonstructural empirical literature can be interpreted in the context of the job search model. We then discuss conditions for identification and methods of estimation. Finally, we describe three empirical papers that have estimated extended versions of the standard model, and thus exemplify the nature of scientific progress in the structural literature, as we discussed in the introduction of this chapter, and we report empirical findings from counterfactual experiments in those papers.

---

[108]Examples of papers based on wage posting models include Eckstein and Wolpin (1990), Kiefer and Neumann (1993), van den Berg and Ridder (1998), Bontemps, Robin and Vandenberg (1999, 2000) and Postel-Vinay and Robin (2002). Those based on search-matching-bargaining .models include, among others, Eckstein and Wolpin (1995), Cahuc, Postel-Vinay and Robin (2006) and Flinn (2006).

1. The Standard Discrete-Time Job Search Model

In the discrete time formulation, an unemployed individual receives a job offer in each period with probability q. Wage offers are drawn from the known cumulative distribution function, $F(w)$. An accepted job offer (and its concomitant wage) is permanent. While unemployed an individual receives b, unemployment benefits (if eligible) net of the cost of search. The individual is assumed to maximize the present discounted value of net income. We consider both infinite and finite horizon models, which have somewhat different empirical implications and implications for the identification of model parameters.[109]

Infinite Horizon Model: The value of a wage offer of w, given a discount factor of $\delta(=\frac{1}{1+r})$, is

$$W(w) \;=\; w + \delta w + \delta^2 w + ... \tag{133}$$

$$\;=\; w + \delta W(w) \text{ or}$$

$$W(w) \;=\; \frac{w}{1-\delta}. \tag{134}$$

In any period, the value of continuing to search , V, either because an offer was rejected or an offer was not received, consists of the current period payoff, b, plus the discounted expected value of waiting another period. In that case, if an offer is received, with probability q, the individual chooses between the maximum of the value of working at a wage w, $W(w)$, or continuing to search and receive V. If no offer is received, which occurs with probability $1 - q$, the individual must continue to search and receives V.Thus the alternative-specific value function, the Bellman equation, for the search choice is

$$V = b + \delta[qE \max(W(w), V) + (1 - q)V]. \tag{135}$$

Rearranging, yields

$$V(1 - \delta) = b + \delta qE \max(W(w) - V, 0), \tag{136}$$

which has a unique solution for $V > 0$, as long as the cost of search is not so large as to make the right hand side negative.[110] Defining w^*, the reservation wage, to be the wage offer that

---

[109]See Mortensen (1986) for the continuous time case.

[110]The LHS is linearly increasing in V and passes through the origin. The RHS is monotonically decreasing in V until it reaches b, and is then constant. There will be a unique intersection, and a unique $V > 0$, as

equates the value of search and the value of accepting the job, that is,

$$w^* = (1 - \delta)V, \tag{137}$$

with a little further algebra, we obtain the following implicit equation for the reservation wage (which must have a unique solution given that V does):

$$w^* = b + \frac{q}{r} \int_{w^*}^{\infty} (w - w^*) dF(w).^{111} \tag{138}$$

Thus, the reservation wage is a function of $b, \frac{q}{r}$ and $F(w)$:

$$w^* = w^*(b, \frac{q}{r}, F)^{112} \tag{139}$$

The individual accepts any wage offer that exceeds the reservation wage and declines offers otherwise.

Although the reservation wage is a deterministic function, the length of an unemployment spell is stochastic because the timing and level of wage offers are probabilistic. Thus, measures of the outcomes of search, such as the duration of unemployment spells and the level of accepted wages, are probabilistic. In particular, the survivor function, the probability that the duration of unemployment is at least as large as some given length, is

$$\Pr(T_u \geq t_u) = \quad [qF(w^*) + (1 - q)]^{t_u} \tag{140}$$

$$= \quad [1 - q(1 - F(w^*))]^{t_u}. \tag{141}$$

The term inside the brackets in (140) is the probability of receiving an offer in a period and rejecting it (because it is below the reservation wag) plus the probability of not receiving on

---

long as $b > -\beta q E \max(W(w), 0)$.

[111] In a continuous time model in which the arrival of offers follows a Poisson process with parameter λ, the implicit reservation wage equation is identical except that the instantaneous arrival rate (λ) replaces the offer probability, q.

[112] Given a distributional assumption for $F(w)$, the solution for the reservation wage involves numerically (if, as is for most distributions the case, there is no closed form solution) solving a non-linear equation. An alternative solution method would be to start from the reservation wage for the final period of a finite horizon problem (see below) and iterate on the reservation wage until it converges. Convergence is assured because the value function is a contraction mapping (see Sargent (1987)).

offer. The cdf, pdf and hazard function are:

$$cdf \quad : \quad \Pr(T_u < t_u) = 1 - [1 - q(1 - F(w^*))]^{t_u}, \tag{142}$$

$$pdf \quad : \quad \Pr(T_u = t_u) = [1 - q(1 - F(w^*))]^{t_u} q(1 - F(w^*)) \tag{143}$$

$$Hazard\ Function \quad : \quad \Pr(T_u = t_u | T_u \geq t_u) = \frac{\Pr(T_u = t_u)}{\Pr(T_u \geq t_u)} = q(1 - F(w^*)) = h. \tag{144}$$

As seen, the survivor function, the cdf and the pdf can all be written as functions of the hazard rate, the exit rate from unemployment conditional on not having previously exited. From (144), it can be seen that the hazard rate is constant. Thus, in a homogeneous population, the infinite horizon search model implies the absence of duration dependence.

Given parameter values, mean duration is given by

$$E(t_u) \quad = \quad \sum_0^\infty t_u \Pr(T_u = t_u) \tag{145}$$

$$= \quad q(1 - F(w^*))^{-1} = \frac{1}{h}.$$

Notice that mean duration is simply one over the hazard rate.[113] Likewise, the mean of the accepted wage is

$$E(w|w \geq w^*) = \int_{w^*}^\infty \frac{w}{1 - F(w^*)} dF(w), \tag{146}$$

which clearly is larger than the mean of the wage offer distribution.[114]

In addition to implying a constant hazard rate, the infinite horizon model has predictions about the impact of changes in b, $\frac{q}{r}$ and $F(w)$ on the reservation wage, on the hazard rate and on the moments of the accepted wage distribution. It thus is, in principle, possible to test the theory. Comparative static effects of the hazard rate (and thus mean duration) with

---

[113]Wolpin (1995) provides a proof.

[114]The distribution of accepted wages is the truncated distribution of wage offers, namely, $f(w|w \geq w^*) = \frac{f(w)}{1 - F(w^*)}$.

respect to its arguments are (see Mortensen (1986)):

$$\frac{dh}{db} = -\lambda f(w^*)\left(\frac{1}{1+h/r}\right) < 0, \tag{147}$$

$$\frac{dh}{dq} = -f(w^*)\left(\frac{w^*-b}{1+h/r}\right) + (1-F(w^*)) \gtrless 0, \tag{148}$$

$$\frac{dh}{d\mu} = -qf(w^*)\left(\frac{h/r}{1+h/r}\right) > 0, \tag{149}$$

$$\frac{dh}{ds} = -qf(w^*)\left(\frac{(\lambda/r)\int_0^{w^*} F_s(w,s)dw}{1+h/r}\right) < 0. \tag{150}$$

An increase in the level of unemployment compensation benefits increases the reservation wage and reduces the unemployment hazard rate (147). An increase in the offer probability has an ambiguous effect; it increases the reservation wage, which reduces the hazard rate, but also directly increases the hazard rate through the higher offer probability (148). It turns out that for a certain class of distributions (log concave), the latter effect dominates (Burdett (1985)). Increasing the mean of the wage offer distribution (149), μ, increases the hazard rate; although an increase in μ increases the reservation wage, the increase is less than one for one. Finally, increasing the mean preserving spread of the distribution 150, s, reduces the hazard because an increase in the mass of the right tail of the wage offer distribution increases the payoff to search, thus increasing the reservation wage. An additional set of implications follow about the mean of the accepted wage; anything that increases the reservation wage also increases mean accepted wage.

Finite Horizon Model: Spells of unemployment tend to be short (weeks or months) in relation to an individual's life span. A finite lifetime would not seem, therefore, to be a reason to explore the implications of a finite horizon search model. On the other hand, it is reasonable to assume that individuals generally will not be able to self-finance extended periods of unemployment and that external borrowing is limited. One can think, then, of the finite horizon as corresponding to the maximal unemployment period that can be financed through internal and external funds, although we continue to assume that once a job is accepted it lasts forever, that is, the horizon is infinite subsequent to accepting a job. In addition to the previous notation, we denote by T the end of the search horizon. To close the

model, it is necessary to specify the value function if the terminal period is reached without having accepted a job. We assume that in that case the individual receives b forever.[115]

Without going into the details, the reservation wage path can be shown to satisfy the following difference equation:

$$\frac{w^*(t)}{1-\delta} = b + \frac{\delta}{1-\delta}w^*(t+1) + \frac{\delta}{1-\delta}q\int_{w^*(t+1)}^{\infty}(w - w^*(t+1))dF(w) \text{ for } t < T, \quad (151)$$

$$w^*(T) = b. \quad (152)$$

Notice that 151 reduces to the implicit reservation wage equation for the infinite horizon problem if $w^*(t) = w^*(t+1)$. Given a distributional assumption for wage offers, $F(w)$, the solution of the finite horizon reservation wage path can be obtained numerically by starting from period T and working backwards.[116]

In the finite horizon case, the reservation wage is decreasing in the duration of the spell, $\frac{dw^*(t)}{dt} < 0$, rather than being constant as in the infinite horizon case. In addition, the reservation wage is bounded from below by b (at T), and from above by the infinite horizon reservation wage (w^*). The hazard rate is thus increasing in t, $\frac{dh(t)}{dt} = -\lambda f(w^*)\frac{dw^*(t)}{dt} > 0$. Thus, the longer the spell duration, the greater the exit rate. The important property of the finite horizon reservation wage is that it depends on the time left until the horizon is reached. The reservation wages are equal under two different horizons not when they have the same amount of time elapsed since beginning the spell of unemployment, but when they have the same amount of time left until the horizon is reached ($T - t$).

Non-Structural (Parametric) Approach to Estimation: The early non-structural approach to estimating the job search model was regression based. The primary concern of that literature, as well as the later literature based on hazard modeling, was the estimation of the impact of unemployment benefits on the duration of unemployment and post-unemployment wages.[117] The regression (or hazard rate) specification in the non-structural approach was

---

[115]An alternative would be to assume that once the terminal period is reached, the individual accepts the next offer that arrives, in which case the reservation wage at T is zero.

[116]See Wolpin (1987), for the particular case in which $F(w)$ is either normal or log normal.

[117]There are statistical issues better handled by specifying the hazard function, such as dealing with incomplete spells and time-varying regressors. See Meyer (1991) for an example of this approach. The issues

97

motivated by the standard job search model. Classen (1977) provides a clear statement of the connection between the theory and the regression specification. The latter is given by

$$D \;=\; \alpha_0 + \alpha_1 WBA + \sum \alpha_i X_i + u_D, \tag{153}$$

$$Y \;=\; \beta_0 + \beta_1 WBA + \sum \beta_i X_i + u_Y \tag{154}$$

where D is spell duration, Y is a measure of the post-unemployment wage, WBA is the weekly UI benefit amount and the $X's$ are "proxies" for a worker's skill level, the cost of search and the job offer rate. As Classen notes, the determinants of both spell duration and the post-unemployment wage should be exactly the same as they are both optimal outcomes derived from the search model. Among the proxy variables used in Classen's analysis are demographics, such as age, race and sex, and a measure of the wage on the job held prior to beginning the unemployment spell. Although not included in the Classen study, other variables often included in this type of specification are education, marital status, number of dependents and a measure of aggregate unemployment in the relevant labor market.

A test of the theory amounts to a test that benefits increase both expected duration and the mean accepted wage, that is, that α_1 and β_1 are both positive. Any further test of the theory would involve specifying how proxy variables are related to the structural parameters, q, the offer probability, and $F(w)$, the wage offer distribution.

Classen is clear as to the purpose of including the pre-unemployment wage, namely as a proxy, most directly perhaps for the mean of the wage offer distribution. However, although the inclusion of that variable or, as is also common, of the replacement rate, the ratio of the benefit level to the pre-unemployment wage, was and continues to be standard in the nonstructural literature, the need for stating a rationale has been lost. As has been pointed out elsewhere (Wolpin (1995)), the inclusion of the pre-unemployment wage (or the replacement rate) cannot be justified by the standard search model, that is, given perfect measures of q and $F(w)$, it would not have any impact on search outcomes. More importantly, however, given its ubiquitous use, is that its inclusion leads to "proxy variable bias," as explained below.

---

we raise, however, are easier to demonstrate in a regression framework, but hold in the hazard framework as well.

Of course, the rationale for its inclusion is to avoid omitted variable bias. For example, suppose that the pre-unemployment wage is meant to proxy the mean of the wage offer distribution faced by individual i, μ_i. Now, UI benefits are usually tied to the pre-unemployment wage, at least up to some limit. Thus, if some determinants of μ_i are omitted (we do not have a perfect measure of $F(w)$), variation in the benefit level will reflect, in part, the fact that those with higher pre-unemployment wages have higher μ_i's. In that case, a positive correlation between UI benefit levels and the pre-unemployment wage will lead to a negative bias in the effect of UI benefits on the duration of unemployment (recall that the higher is μ_i, the greater the hazard rate).

Although omitted variable bias is well understood, the potential for bias introduced by using proxy variables is less well appreciated. The source of the problem is that, in the context of a search model, the pre-unemployment wage must have been the outcome of a search during a prior unemployment spell. To isolate the effect of using this proxy, assume that the duration of the prior unemployment spell was governed by the same behavioral process and fundamentals as the current spell. Such an assumption is consistent with a model in which there are exogenous layoffs and unemployment spells are renewal processes (the stochastic properties of all unemployment spells are the same).[118] In particular, suppose that only the benefit level and the mean of the wage offer distribution vary (say, geographically) in the sample. Then, taking deviations from means (without renaming the variables, to conserve on notation), the duration equation is

$$D_i = \pi_{11} b_i + \pi_{21} \mu_i + v_{1i}. \tag{155}$$

Assuming μ_i is unobserved, and thus omitted from the regression, the bias in the ols estimator of π_1 is given by

$$E(\widehat{\pi}_{11} - \pi_{11}) = \pi_{21} \frac{\sigma_\mu^2}{\sigma_b^2} \beta_{b,\mu}, \tag{156}$$

where $\beta_{b,\mu}$ is the regression coefficient of b on μ. Thus, if b and μ are positively correlated and $\pi_{21} < 0$ as theory suggests, the bias in the estimated effect of UI benefits on duration will be negative.

---

[118] The assumption that new unemployment spells are renewal processes rules out any structural connection between spells, for example, it rules out that the benefit level depends on the pre-unemployment wage.

Now, because the pre-unemployment wage, Y_{-1}, is the result of a prior search, it will have the same arguments as (155), namely

$$Y_{i,-1} = \pi_{12}b_i + \pi_{22}\mu_i + v_{2i} \tag{157}$$

$$= \pi_{22}\mu_i + \omega_i, \tag{158}$$

where $\omega_i = \pi_{12}b_i + v_{2i}$ and where $E(\omega b_i) \neq 0$ given the definition of ω_i. To derive a regression equation that includes b_i and $Y_{i,-1}$, solve for μ_i in (158) and substitute into (155), yielding

$$D_i = \pi_{11}b_i + \pi_{21}\frac{(Y_{i,-1} - \omega_i)}{\pi_{22}} + v_{1i} \tag{159}$$

$$= \pi_{11}b_i + \frac{\pi_{21}}{\pi_{22}}Y_{i,-1} + v'_{1i}. \tag{160}$$

where $v'_{1i} = -\frac{\pi_{21}}{\pi_{22}}\omega_i + v_{1i}$. We are interested in whether the ols estimator for π_{11} is biased, that is, whether $E(b_i v'_{1i}|w^a_{i,-1}) = E(-\frac{\pi_{21}\pi_{12}}{\pi_{22}}b_i^2|\pi_{22}\mu_i + \pi_{12}b_i + v_{2i} = w^a_{i,-1}) = 0$. It is easiest to see whether this holds by explicitly deriving the bias expression. It is given by

$$E(\widehat{\widehat{\pi}}_{11} - \pi_{11}) = \pi_{21}^2 \frac{\sigma_\omega^2 \sigma_\mu^2}{\sigma_b^2 \sigma_{w^a_{-1}}^2 - (\sigma_{bw^a_{-1}})^2}(\pi_{21}^{-1}\beta_{b,\mu} - \pi_{22}\beta_{b,\omega}). \tag{161}$$

The bias is zero if either $\sigma_\omega^2 = \pi_{12}^2\sigma_b^2 + \sigma_{v_2}^2 = 0$ or if $\pi_{21}^{-1}\beta_{b,\mu} - \pi_{22}\beta_{b,\omega} = 0$, either of which would be a fortuitous property of the sample.

What is the relationship between the biases from omitting μ_i (omitted variable bias) versus including $w^a_{i,-1}$ (proxy variable bias)? It turns out that a sufficient condition for the bias from omitting the pre-unemployment wage to be smaller than from including it is that $\beta_{b,\omega} = \pi_{12}\sigma_b^2 = 0$, which only will hold if $\pi_{12} = 0$, a violation of the theory. In general, the biases cannot be ordered and it is unclear which is the better strategy to follow to minimize the bias. The implicit (that is, without justification) assumption made by almost all researchers is that omitted variable bias is greater than proxy variable bias in this context. Moreover, if the benefit level varies with the pre-unemployment wage and we have good measures of the mean of the wage offer distribution, variation in the benefit level from this source is helpful in identifying the UI benefit effect. The variation in the pre-unemployment wage around the mean of the offer distribution that induces benefit variation is purely due to random draws from the wage offer distribution.

Researchers adopting the nonstructural approach have universally included the pre-unemployment wage, and thus, assumed that omitted variable bias is greater than proxy variable bias. There is a larger point reflected by this choice, namely the importance of theory in empirical work. Structural work requires that all variables be explicitly accounted for in the model. A similar standard for nonstructural work might have revealed the existence of the choice between omitted variable and proxy variable bias, a choice not explicitly acknowledged in the nonstructural literature.[119]

Structural Approach to Estimation:

Identification   Identification is no less an important issue in structural empirical work than in nonstructural work. We consider identification of the standard search model parameters assuming we have data from a homogeneous population on durations of unemployment and on accepted wages.[120]

To establish identification, it is useful to rewrite the reservation wage implicit equation (138) as

$$
w^* = b + \frac{q}{r} \int_{w^*}^{\infty} (w - w^*) dF(x) \tag{162}
$$

$$
= b + \frac{q}{r} [E(w|w \geq w^*)(1 - F(w^*)) - w^*(1 - F(w^*)) \tag{163}
$$

$$
= b + \frac{h}{r} [E(w|w \geq w^*) - w^*], \tag{164}
$$

where the last equality uses (144). From the accepted wage data, note that a consistent estimator of the reservation wage is the lowest observed wage: $\lim_{n \to \infty} \Pr(|w_{\min} - w^*|) = 0$.[121] Then, given an estimate of h from the duration data, and recognizing that from the accepted

---

[119] As we noted, it is also usual to include some aggregate labor market statistic like the local unemployment rate. The idea is that the aggregate statistic reflects labor market demand and so will affect the offer rate or the wage offer distribution. As shown in Wolpin (1995), because the aggregate statistic is simply the aggregation of the search decisions over the unemployed population, it does not reflect solely demand conditions and estimates of UI benefit effects suffer from proxy variable bias.

[120] This section follows the development in Flinn and Heckman (1981) and Wolpin (1995).

[121] The minimum observed wage is a superconsistent estimator of the reservation wage in that it converges at rate N. This leads to non-standard asymptotics in the likelihood estimation of the search model (see Flinn and Heckman (1981) and Christensen and Kiefer (1994))..

wage data, we can also obtain an estimate of $E(w|w \geq w^*)$, we can identify b (which includes the unobserved cost of search) if we take r as given. This result does not require a distributional assumption for wage offers.

We cannot, however, separate q and F without a distributional assumption. Although we know $F(w|w \geq w^*)$, given an estimate of w^*, we can recover the wage offer distribution, $F(w)$, only if it is possible to recover the untruncated distribution from the truncated distribution. Obviously, that cannot be done without making a distributional assumption. Assuming that $F(w)$ is recoverable from the accepted wage distribution, then from $h = q(1 - F(w^*))$, we can recover the offer probability, q.

Recoverability of the wage offer distribution is not always possible. It is useful for later reference to consider an example taken from Flinn and Heckman (1982). Assume that wage offer distribution is Pareto, that is, having pdf

$$f(w) = \varphi w^\gamma, \ \varphi = \frac{-(\gamma+1)}{c^{\gamma+1}} \ c \leq w \leq \infty.. \tag{165}$$

Notice that the support of the distribution is bounded from below by a constant c. The density of accepted wages is

$$f(w|w \geq w^*) = \frac{\varphi w^\gamma}{1 - F(w^*)} \tag{166}$$

$$= \frac{-(\gamma+1)}{(w^*)^{\gamma+1}} w^\gamma, \ w \geq w^* \geq c. \tag{167}$$

Given an estimate of w^*, from the minimum observed wage, and of the conditional density, we can recover γ. However, there are many values of c that are consistent with the truncated distribution. We, thus, cannot identify φ, which means we cannot identify the wage offer distribution, F. As already noted, the consequence of this lack of identification is that we cannot separate q and F, or, more specifically, φ. To see that explicitly, write the reservation wage equation and the hazard function under the Pareto distribution,

$$w^* = b + \frac{q}{r} \int_{w^*}^\infty (w - w^*) dF(x) \tag{168}$$

$$= b + \frac{q\varphi}{r} \left(\frac{(w^*)^{\gamma+2}}{r(\gamma+1)(\gamma+2)} \right);$$

$$h = \frac{-\lambda\varphi}{\gamma+1} (w^*)^{\gamma+1}. \tag{169}$$

102

Because q and φ only enter multiplicatively as $q\varphi$, it is impossible to separately identify them. Fortunately, most of the commonly used distributions for wage offer functions, for example, the lognormal distribution, are recoverable from the distribution of accepted wages. However, there is an important lesson to draw, namely that parametric assumptions do not always assure identification.

The analysis of identification in the finite horizon case is similar. The reservation wage at each period can be consistently estimated from the period-specific minimum observed wages. Analogous to (164), we can write the implicit reservation wage equation as

$$w_t^* = b(1-\delta) + \delta w_{t+1}^* + \delta h_{t+1}[E(w|w \geq w_{t+1}^*) - w_{t+1}^*].^{122} \tag{170}$$

This equation must hold exactly at each time t. As long as there are durations of unemployment that extend through more than two periods, that is, given $w_t^*, w_{t+1}^*, w_{t+2}^*, h_{t+1},$ $h_{t+2}, E(w|w \geq w_{t+1}^*)$ and $E(w|w \geq w_{t+2}^*)$, δ (or r) and b can be separately identified by solving the two difference equations for the two unknowns, δ and b. Recall that this separation was impossible in the infinite horizon case. Moreover, if we have more than three periods of data, the model is rejectable. Identifiability of q and F, however, still requires recoverability of the wage offer distribution.

Likelihood Function The likelihood function for the search model takes an analogous form as that for the binary labor force participation model, given data on unemployment durations and accepted wages. Consider the likelihood contribution of an individual solving an infinite horizon search model, who has a completed unemployment spell of length d_i and accepted wage w_i^a :

$$
\begin{aligned}
L_i &= [q\Pr(w < w^*) + (1-q)]^{d_i - 1} \times q\Pr(w_i^a > w^*, w_i^a) \tag{171}\\
&= [q\Pr(w < w^*) + (1-q)]^{d_i - 1} \times q\Pr(w_i^a > w^*|w_i^a)f(w_i^a). \tag{172}
\end{aligned}
$$

Note that the reservation wage, the solution to the implicit equation (138), is a function of the model parameters b, δ, q and $F(w)$. The first (bracketed) term in the likelihood is the probability that in each of the periods up to $d_i - 1$ the individual received an offer and rejected

---

[122]Setting $w_t^* = w_{t+1}^* = w^*$ and $h_{t+1} = h$ yields the infinite horizon implicit reservation wage function.

it or did not receive an offer. The second term is the probability that the individual received an offer of w_i^a in period d_i and accepted it. Individuals who have incomplete unemployment spells would contribute only the first term to the likelihood.

In the labor force participation model of section III.A.1, the probability of accepting an offer to work depended on both a random wage draw and a random taste draw. Thus, whether or not an individual worked conditional on a wage draw was probabilistic, because it depended on the taste draw that we do not observe. In the search model, however, for any given value of the reservation wage (or the parameters that determine it), whether or not an individual works conditional on a wage draw is deterministic, that is, its probability, $\Pr(w_i^a > w^* | w_i^a)$ in (172), is either one or zero. In order that the likelihood not be degenerate, the reservation wage must be less than the lowest accepted wage in the sample; for this reason, the minimum observed wage in the sample is the maximum likelihood estimate of the reservation wage. That would not create an issue if observed wages were all reasonable. However, in most survey data sets the lowest reported wage is often quite small, less than one dollar or even only a few cents. Such outliers would potentially have an extreme effect on the estimates of the structural parameters. One remedy would be to be to trim the wage data, say by whatever percent led to a "reasonable" lowest wage. However, that would be essentially choosing the reservation wage by fiat. A second alternative would be to add another error to the model, for example, by allowing the cost of search to be stochastic in which case the reservation wage would be stochastic.[123] A very low accepted wage would be consistent with the individual having drawn a very high search cost.

Of course, adding another source of error does not deal with what is the likely root cause of the problem, which is that wages are not accurately reported.[124] That "fact" has led researchers to directly allow for measurement error in the reported wage. Introducing

---

[123]The implicit reservation wage equation (138) would no longer hold in this case. In particular, the integration would have be taken also over the distribution of the unobserved cost of search recognizing that the reservation wage would be a function of that cost.

[124]In some instances, wage rates that are directly reported in hours or weeks are inaccurate. In other cases, wage rates are derived from a division of earnings, reported over a longer period, say annually, and hours worked reported over that period. The inaccuracy arises from a seeming mismatch in the time period between earnings and hours.

measurement error not only accounts for a real feature of the data, it is also convenient in that it does not require any change in the solution of the search problem. The reservation wage is itself unaffected by the existence of measurement error; the implicit reservation wage equation is still given by (138). Taking into account the existence of measurement error, letting w_i^{aR} be the reported accepted wage and w_i^{aT} the true accepted wage, the likelihood function (172) becomes

$$
\begin{aligned}
L_i &= \left[q \Pr(w_i^{aT} < w^*) + (1-q) \right]^{d_i - 1} \times q \int_{w_i^{aT}} \Pr(w_i^{aT} > w^*, w_i^{aR}) \qquad (173) \\
&= \left[q \Pr(w_i^{aT} < w^*) + (1-q) \right]^{d_i - 1} \times q \int_{w_i^{aT}} \Pr(w_i^{aT} > w^* | w_i^{aR}, w_i^{aT}) g(w_i^{aR} | w_i^{aT}) dF(w) \\
&= \left[q \Pr(w_i^{aT} < w^*) + (1-q) \right]^{d_i - 1} \times q \int_{w_i^{aT}} \Pr(w_i^{aT} > w^* | w_i^{aT}) g(w_i^{aR} | w_i^{aT}) dF(w),
\end{aligned}
$$

where $g(w_i^{aR} | w_i^{aT})$ is the distribution of the measurement error and where the third equality emphasizes the fact that it is the true wage only and not the reported wage that affects the acceptance probability. The most common assumption in the literature is that the measurement error is multiplicative, that is, proportional to the true wage.

The estimation of the finite horizon search problem when there are extreme low-wage outliers is particularly problematic. Recall that the reservation wage is declining with duration. Thus, if an outlier observation occurs at an early duration, the entire subsequent path of reservation wages must lie below the reservation wage at that early duration. The incorporation of measurement error is, therefore, critical for estimation. The analogous likelihood contribution for an individual for the finite horizon model, which takes into account that the reservation wage is duration dependent, is

$$
L_i = \prod_{j=1}^{d_i - 1} \left[q \Pr(w_j^{aT} < w_j^*) + (1 - q_j) \right] \times q \Pr(w_{d_i}^{aT} > w_d^*, w_d^{aR}) \qquad (174)
$$

The estimation of the partial equilibrium search model involves the same iterative process as for the labor force participation model, namely numerically solving a dynamic programming problem at trial parameters and maximizing the likelihood function. The generality

of the DCDP approach has allowed researchers considerable flexibility in modeling choices. Thus, researchers have adopted different distributional assumptions and have extended the standard search model in a number of directions that we have already mentioned.

As has been generally true in the DCDP literature, the theoretical models that serve as their foundation cannot directly be taken to the data. In the case of the standard search model, neither the infinite horizon nor the finite horizon model can fit the generally observed fact that the hazard rate out of unemployment declines with duration. Recall that the hazard rate is constant in the infinite horizon case and increasing in the finite horizon case. There are several ways to deal with this mismatch between the data and the models. In the infinite horizon model, introducing unobserved herterogeneity in model fundamentals, such as the cost of search, the offer probability and/or the wage offer distribution, can produce negative duration dependence in the population hazard while maintaining stationarity at the individual level. In the finite horizon case, allowing for time dependencies in model fundamentals, such as allowing offer probabilities to decline with duration, can create negative duration dependence, in this case at the individual level as well as at the population level.

Selected Literature: Rather than do a comprehensive review of the contributions of DCDP modeling to the empirical (partial equilibrium) search literature, we illustrate the broad range of model specifications and the usefulness of the approach for policy evaluation with three examples.

Rendon (1997, 2006) In this first example, Rendon (1997, 2006) extends the standard finite horizon search model to allow for exogenous job loss (layoffs), for on-the-job search and for savings in the presence of borrowing constraints. Recall that the standard finite horizon search model imposes a terminal search period, with the putative rationale being that the individual cannot search indefinitely due to a limit on borrowing. However, because this limitation is not an explicit part of the model, the terminal value function (or equivalently, the terminal reservation wage) is not determined as part of the model. Rendon (1997, 2006) fills this lacuna in the structural empirical literature.[125]

---

[125]Theoretical models in which job searchers faced borrowing constraints appeared much earlier, starting with Danforth (1979). However, formal empirical implementation did not become feasible until the develop-

Rendon, building on theoretical papers by Danforth (1979) and others and on the DCDP model of Wolpin (1992), considers a job search model with the following features:

1. Individuals maximize the present discounted value of lifetime utility. Flow utility is a CRRA function in consumption. Individuals are finitely lived and exogenously retire at a known time prior to the end of life. Time is discrete. Prior to retirement, the individual is in one of two employment states, unemployed or employed.

2. As in the standard search model, in each period of unemployment, the individual receives a job (wage) offer with a positive probability. If an offer is received, the individual makes an acceptance-rejection decision. The individual enters each period with some level of assets (positive or negative) and decides the level of assets to carry forward to the next period. Income while unemployed consists of the interest return (or payment, if assets are negative) on assets, and unemployment compensation benefits plus other family transfers minus the cost of search.

3. In each employment period, the individual faces a positive probability of layoff as well as a positive probability of receiving a wage offer from another employer. Regardless of whether an offer is received, the individual can decide to quit and become unemployed. As in the unemployment state, the individual decides on the level of assets to carry forward into the next period. Income while employed consists of the interest return (or payment, if assets are negative) on current assets plus the wage. Wages grow, starting at the initial accepted wage, deterministically with job tenure.

4. The level of assets that an individual holds in any period can be negative; the individual can carry debt, but the amount of debt cannot fall below the present value of the amount the individual can pay back with certainty (the Hakansson-Miller or "natural" borrowing limit). Because the individual can with some small probability be unemployed until the retirement date, the only certain income each period is the amount of unemployment income given by UI benefits and family transfers net of search costs.

In this model, individuals generally accumulate assets while employed as insurance against future unemployment spells and decumulate assets to finance search while unemployed. An individual's reservation wage, as in the standard model, declines with duration as assets are

ment of estimation methods for DCDP models.

run down. Individuals who start an unemployment spell (at the same life cycle point, say, due to a layoff) with higher assets (having randomly drawn a higher acceptable wage offer on the prior spell of unemployment) have higher reservation wages, longer unemployment spells and higher accepted wages. Thus, wages will not only be positively correlated across employment spells generated by job-to-job transitions, but also across unemployment spells separated by layoffs.

Unemployment spells are not only generated by exogenous layoffs, but also by voluntary quits into unemployment, even though wages on the job are non-stochastic. This behavior can arise when the offer probability is greater while unemployed than while employed. Consider an unemployed individual who, having not received frequent offers or only received offers at low wages, has drawn down assets to finance consumption, perhaps even hitting the borrowing constraint. The individual in this situation optimally accepts a low wage job. The individual, once employed, will begin to accumulate assets as insurance. At some time, the individual, having not received any higher wage offer from another firm will have accumulated sufficient assets for it to be optimal to quit into unemployment, financing consumption with the accumulated assets, to take advantage of the higher job offer rate during unemployment.[126]

As seen by this discussion, the existence of voluntary quits requires a certain parameter configuration. Given data in which voluntary quits arise, this parameter configuration must be an outcome of the estimation. Most DCDP models, like this search model, have the characteristic that model predictions are parameter dependent. This characteristic does not imply that these models do not have rejectable restrictions. Indeed, DCDP models are generally highly restrictive. Recall that the standard finite horizon model was rejectable, not in the conventional way of testing comparative static predictions, but because only a few parameters determined the entire profile of reservation wages. Tests of DCDP models are best thought of as tests arising from cross-equation restrictions. Models like the one estimated in Rendon also have such cross-equation restrictions, but they do not have easily

---

[126]The search model does not have to incorporate savings for it to be optimal for individuals to quit into unemployment It is sufficient that there be a finite horizon (retirement) and either that the offer rate be higher in unemployment or that there be a wage return to general experience (Wolpin (1992)).

derived analytical forms. However, to the extent that those cross-equation restrictions are seriously violated, the model will not be able to fit the data very well. Tests of model fit are (imperfect) tests of the model's implicit cross equation restrictions.

In addition to the discrete state variable, job tenure, the search problem in Rendon has two continuous state variables, assets and the accepted wage. The latter is a state variable because the accepted wage is permanent (and wage growth is deterministic), which implies that reservation wage for accepting an offer from another employer depends on the wage at the current employer. Rendon solves the DP model by discretizing assets and wages, a method we discussed above. However, given the fine discretization he used, it was not tractable to solve and estimate the model over his postulated 40.5-year working life on a quarterly basis, that is for 162 quarters. To make it tractable, Rendon solved the model on a quarterly basis for the first 12.5 years, then on an annual basis for the next 8 years and finally a biannual basis for the next 20 years.[127]

Rendon's model is extremely parsimonious. It contains only12 parameters that must account for all of the labor force transitions, wages and assets observations of a sample of white male high school graduates over the first 40 quarterly subsequent to their graduation.[128] The model is estimated by maximum likelihood. As is conventional in the DCDP literature, Rendon computes chi-square statistics for the match between the actual data and data generated by the model estimates for a wide range of statistics. The quality of the fit of the model is mixed.

The estimated model is used to perform a number of counterfactual exercises. In particular, Rendon considers the impact on labor market outcomes of relaxing the extent to which borrowing constraints are binding. In the quantitative experiment he performs, he finds that allowing agents to borrow up to 50 percent of the natural borrowing limit, as opposed to the estimated 10 percent in the baseline, would increase the duration of unemployment in the first period after graduating from high school by 12.5 percent and increase the accepted wage on that first job by one-third. Given a greater ability to borrow to finance unemployment spells, agents will hold fewer assets throughout their life cycle; in the experiment, asset hold-

---

[127]This procedure for reducing the size of the state space follows Wolpin (1992).

[128]Assets are only observed annually. The data are from the NLSY79.

ings would be one-third less 10 years after graduation. Thus, Rendon finds that borrowing constraints importantly affect employment outcomes and asset accumulation.

Paserman (2008) The standard search model is based on the conventional assumption that agents use exponential discounting in weighing the current net cost of search and the future wage payoff from continuing to search. Paserman (2008), following a growing literature in which agents are assumed to have time-inconsistent preferences, specifies and estimates a search model with (quasi) hyperbolic discounting. In addition to allowing for present-biased preferences, Paserman extends the search model to include (i) a decision about search intensity, in essence, a choice about the per-period probability of receiving a job offer, (ii) an exogenous probability of layoff once employed and (iii) the receipt of unemployment benefits for a fixed period of time. The agent solves a finite horizon problem until the point at which unemployment benefits are exhausted and an infinite horizon problem from that point forward. Thus, the reservation wage and search intensity are constant after exhaustion, but are duration dependent during the period when the agent is still eligible for unemployment benefits.

To see the role of hyperbolic discounting, consider the simple discrete time finite horizon search model, abstracting from the additional extensions introduced by Paserman. With hyperbolic discounting, the value functions are:

$$
\begin{aligned}
V_t &= b + \beta\delta[qE\max(W(w), V_{t+1}) + (1-q)V_{t+1}], \qquad (175)\\
W(w) &= \frac{w}{1-\delta},
\end{aligned}
$$

where as before V_t is the value of searching in period t, $W(w)$ is the value of accepting a wage of w, b is unemployment income net of the cost of search in period t, q is the offer probability, δ is the "long-run" discount factor and β is the "short-run" discount factor. Note that the value of accepting a wage at $t+1$ after searching in period t, as viewed at t, is discounted by δ, that is, exponentially. Thus, it is as if the agent has two selves. The agent who is making a decision in the current period, the current self, is impatient, discounting the expected future payoff to search by $\beta\delta$. However the future self, the self who will receive the benefit of the search and controls future decisions, discounts exponentially. In formulating (175), it was

assumed that the current self is aware that when the next period is reached, the current self at that time will be impatient, in which case the agent is deemed sophisticated. This is in contrast to a naive agent, who instead would assume that in the next period the current self would no longer be impatient.

The reservation wage is, as before, the wage that equates the value of search and the value of employment, namely $w_t^* = (1 - \delta)V_t$. With a little algebra, we can write (175), analogous to (170), as

$$w_t^* = b(1 - \delta) + \beta\delta w_{t+1}^* + \beta\delta h_{t+1}[E(w|w \geq w_{t+1}^*) - w_{t+1}^*]. \tag{176}$$

Obviously, the reservation wage equation with hyperbolic discounting is the same as that with exponential discounting if $\beta = 1$. Further, given its recursive structure, it is clear also that the reservation wage path is lower at all t as the degree of impatience is greater, that is, as β is smaller. Thus, the effect of present-bias in agent preferences is to make job acceptance occur sooner, leading to shorter durations of unemployment and lower accepted wages. The future self, however, would have preferred that the current self be more patient.

As noted, Paserman's model is somewhat more complicated. In particular, he allows for a choice of search intensity, which affects the probability of receiving an offer and which is costly. In this setting, an agent has two instruments to minimize the current cost of search, the choice of search intensity and the choice of the acceptance wage. DellaVigna and Paserman (2006) show that with hyperbolic discounting (in an infinite horizon setting) agents will choose a lower search intensity and a lower reservation wage. Because a lower search intensity leads to a lower offer probability, and thus to longer spells, while a lower reservation wage leads to shorter spells, as in the case of the standard search model, the net effect on expected spell duration is, in general, ambiguous. However, DellaVigna and Paserman show that, as in the standard model, log concavity of the wage offer distribution is sufficient for the expected duration of unemployment to fall.

As in the standard finite horizon model, we can use (176) to consider identification. In the standard model, we noted that b and δ could be separately identified with data on at least three periods that include accepted wages. However, identification is no longer possible with the addition of hyperbolic discounting; one cannot separately identify the structural parameters b, δ and β from $b(1 - \delta)$ and $\beta\delta$. But, Paserman also has data on UI benefits.

In his model, b is a composite of the level of UI benefits, b^{UI}, and the value of search time, b_0.[129] Thus the first term in (176) is $(b^{UI} + b_0)(1 - \delta)$. Clearly, cross-section variation in b^{UI} is sufficient to identify δ, which implies that b_0 and β are also identified.[130]

The model is estimated by maximum likelihood, with an extended version of likelihood function (174) to account for unobserved heterogeneity and layoffs.[131] The estimation method, like all DCDP models, requires iterating between the solution of the DP problem and calculation of the likelihood function. As is standard in the DCDP literature, Paserman presents evidence on goodness-of-fit to evaluate the performance of the model.

The long-run discount factor, when not estimated to be at the boundary, could not be distinguished from one. The short-run discount factor was estimated to be .40 for a low-skilled sample, .48 for a medium-skilled sample and .89 for a high skilled sample. The p-value for a likelihood ratio test of whether the estimates of β were equal to one were less than 0.01 for the first two samples and 0.08 for the third.

Paserman uses the estimates of the model to assess the impact of policy interventions on unemployment search outcomes and on welfare. Measuring welfare in a hyperbolic discounting model is, however, somewhat problematic as there are, in essence, two agents (selves). Paserman adopts as the welfare measure the exponentially discounted utility of the long-run self under the strategy chosen by the hyperbolic self. In an exponential discounting setting, because there is a single agent making optimal choices subject to constraints, any additional constraints must always reduce welfare. With hyperbolic discounting, this is not necessarily the case. Using this welfare measure, Paserman addresses the question of whether it is possible to design a policy that not only improves welfare, but also reduces unemployment duration and lowers government outlays.

Paserman finds that by imposing a fine, equal to the amount of unemployment benefits,

---

[129]Note that the cost of search in his model is isomorphic to the probability of receiving an offer.

[130]The resevation wage, mean accepted wage and hazard rate are all functions of b^{UI}. They can be estimated as nonparametric functions of b^{UI} Paserman also models search after the exhaustion of benefits. In that case, it is assumed that individuals solve an infinite horizon problem.

[131]The model was estimated both for a normal and log normal wage offer distribution. As found by others (for example,Wolpin (1987)), the normal distribution assumption proved problematic. Paserman also allows for measurement error for the reasons previously discussed.

on unemployed agents who do not meet a search effort threshold, it is possible to achieve all of these goals. Indeed, Paserman shows (numerically) that there exists a threshold level of search intensity at which unemployment durations fall and for which the increase in agent welfare and the savings in government outlays is maximized. This experiment implies that program in which the search intensity of unemployed workers is monitored not only may reduce the cost of the UI system (subject to the cost of monitoring), but may also improve the welfare of those who are unemployed.

Ferrall (1997)   As we have stressed, DCDP models have been used extensively for policy evaluation. In the present context, for example, most empirical applications of the DCDP approach to job search provide a quantitative assessment of the impact of changes in the UI system, such as altering benefits. Many of those models capture some, but not all, of the features of the UI system, often in a somewhat stylized fashion. It is reasonable to suppose that the closer a model mimics UI program rules, the better the model will be in evaluating policy changes. Ferrall (1997) structurally estimates a DCDP model of job search, which integrates all of the major features of the Canadian UI system.[132]

Ferrall studies the transition from school to work. In Canada, as in the U.S., the period of search for one's first job after leaving school is usually not covered by the UI system. Although that spell of search unemployment is not insured, there is still the potential for the UI system, given its structure, to affect search behavior. To understand why, consider the structure of the UI system in Canada relevant during the time period studied by Ferrall. In that system:

1. An unemployed worker who is eligible for insurance must wait 2 weeks after becoming unemployed before collecting benefits.[133]

---

[132] Ferrall also estimates a model for the U.S., but, because the UI system varies from state to state and is much less generous than the Canadian system, he does not incorporate UI benefits into the analysis. We focus on the Canadian data to highlight the fact that the DCDP approach allows for a detailed representaion of UI policy.

[133] Actually, this waiting period is only for those who are unemployed through layoff. Those who quit or were fired had a waiting period of five weeks. Ferrall assumes the waiting period to be two weeks independent of the reason for the unemployment spell.

2. The benefit level depends on the previous wage through a fixed replacement rate (.60 at the time) The insurable weekly wage is bounded from below by \$106 and from above by \$530. Thus benefits are \$0 if the wage on the previously held job was less that \$106, .6 times the wage if the wage is between the bounds and \$318 for wages at or above \$530.

3. To be eligible for UI benefits, a person must have worked on insurable jobs a certain number of weeks in the 52 weeks prior to becoming unemployed. The number of weeks depends on the regional unemployment rate and the person's previous employment and UI history.

4. The number of weeks of benefits depends on the number of weeks worked on the previous job and on whether the individual is eligible for extended benefits, but is capped at 52.

There are two elements of the UI system that would affect search during an uninsured spell. First, because benefits are paid during insured spells, there is an incentive for individuals in an uninsured spell to become employed to be eligible for benefits during future unemployment spells. Thus the UI system reduces the reservation wage in an uninsured spell; further, the reservation wage will be lower the higher are benefits (Mortensen (1977)). On the other hand, because the level of benefits increases with the wage on the prior job, there will be an incentive for someone in an uninsured spell to wait for a higher wage offer, that is, to have a higher reservation wage. This incentive, however, only applies to individuals whose reservation wage would otherwise be below the maximum insurable wage.[134] Moreover, the magnitude of these incentive effects depend on the probability that an individual will be laid off from future jobs.

The model estimated by Ferrall, aside from the explicit incorporation of UI rules, differs from the standard single spell search model in a number of ways. The model allows for a search period during school, an initial uninsured spell after leaving school, the first job spell and a subsequent insured unemployment spell if a layoff occurs. The individual maximizes the expected present value of the log of consumption, where consumption equals the wage

---

[134]Recall that inclusion of the wage on the previous job as a proxy for omitted determinants of the duration of a covered unemployment spell is universal in the nonstructural literature. Accounting for the fact that benefits depend on the prior wage provides a rationale for that specification.

while working and the sum of UI benefits net of the cost of search plus the value of home production. In each period of unemployment, the individual receives a job offer with some positive probability. However, a job offer comes not only with a wage offer, but also with a layoff rate. The wage offer function is Pareto and the layoff rate can take on a fixed number of values, randomly drawn.[135] Individuals differ, according to their unobserved type, in their market skill level and in their value of home production.

The solution method is by backwards recursion, where the value function for the infinite horizon search problem when UI benefits are exhausted after a layoff occurs serves as the terminal value function for the insured unemployment spell at the time of benefit exhaustion. The model is solved backwards from there as a finite horizon problem until the beginning of the search period while in school. The estimation is by maximum likelihood. Ferrall provides evidence of model fit.

Ferrall performs a number of counterfactual experiments that vary the parameters of the UI system. The most extreme is the elimination of the UI system, an out-of-sample extrapolation that is only possible within the structural framework. The resulting impact on unemployment durations depends on geographic location and education.[136] Recall from the earlier discussion that there was no unambiguous prediction of how reservation wages would be affected by such an experiment. Ferrall finds that for those with at most a high school education residing outside of the Atlantic region, reservation wages rise; the expected duration of unemployment after leaving school (including those who have no unemployment spell) is estimated to increase by about 50 percent. Similarly, for those with some college residing outside of the Atlantic region, the increase is 40 percent. However, there is almost no effect on the expected duration for those residing in the Atlantic region regardless of education.

---

[135] As previously noted, the lower bound of the support for the Pareto distribution cannot be identified. Ferrall fixes that value. Christensen and Kiefer (2009) also use the Pareto distribution and impose, based on the wage posting model of Burdett and Mortensen (1998), the individual's reservation wage as the lower bound.

[136] UI system parameters depend on region and some of the structural parameters are allowed to differ by education.

C. Dynamic Models of Schooling and Occupational Choices

This section describes the use of DCDP models in labor economics to study schooling and occupational choice and to analyze the effects of policy interventions aimed at increasing skill investment, such as tuition and school attendance subsidies and student loan programs. We begin with a brief discussion of the foundational schooling and occupational choice models from the early literature, which tended to be either static models or life-cycle models without uncertainty. These first generation models were influential in shaping the questions addressed and models developed in the later DCDP literature. We then describe contributions to the more modern DCDP literature.

1. Foundational Literature

One of the earliest discussions of the determinants of schooling and occupational choices is given by Walsh (1935), which describes a model in which individuals invest in education until the return on education equals the return on other possible investments.[137] The paper also examines the empirical support for the model using data from a variety of sources.[138] Walsh (1935) calculates the returns associated with different levels of schooling and with a subset of occupations (doctor, lawyer, engineer), adjusting for costs (tuition, room and board) and foregone earnings. He raises the potential problem of ability bias in comparing lifetime earnings streams of different education levels and different professions. After finding that the wage returns to being a college graduate and to being a lawyer greatly exceed costs, whereas the returns to receiving an M.A. and PhD. are lower than the cost, he attributes the difference to nonpecuniary benefits associated with working in academia.

Roy (1951), in a seminal paper, provides the modern framework for modeling occupational choice as an earnings maximization problem that he uses to analyze the implications of self-selection into occupations for earnings distributions. The Roy model assumes that individuals

---

[137] Walsh (1935) is cited by Becker (1964), perhaps the most influential work in the development of the human capital literature, along with Friedman and Kuznets (1954), as "brilliant work."

[138] The sources include a survey of 15,000 former members of the Alpha Kappa Psi fraternity on the education and earnings, a survey of Land Grant colleges made by the U.S. Department of Interior, published figures on the earnings of physician and doctors, and a survey of Harvard Law graduates.

are endowed with two different skills, drawn from a joint log normal distribution. Each skill is productive in only one of two occupations, denoted by Roy as hunting and fishing. Skill is measured in units of output produced. Thus, an individual's earnings in an occupation are the product of the price of a unit of occupation-specific output times the amount of skill (output production) embodied in the person; individuals choose to work in the occupation that maximizes their earnings. Roy (1951) did not apply the model to data, but showed that the structural parameters of the underlying model, the means, variances and covariances of the joint skill distribution, can be recovered from earnings data, even though earnings in a particular occupation is only observed for people who chose that occupation. The identification of these structural parameters derives from the theoretical formulation of the determination of earnings and from the distributional assumption.[139]

The literature started by Roy (1951) emphasized the importance of self-selection, skill heterogeneity and latent skills in understanding occupational choices and earnings. However, in Roy's formulation, skills are treated as endowments. Another branch of the literature, associated with Mincer (1958), Becker (1964, 1967) and Ben-Porath (1967), evolved with the aim of understanding the human capital investment (or skill acquisition) decision and the implications for lifetime earnings of acquiring skills through schooling and job training investments. Mincer (1958) proposes a lifetime earnings model where the only cost of schooling is foregone earnings. In his model, all individuals are assumed to be *ex ante* identical, which implies a compensating earnings differential for individuals who spend more time in school. In equilibrium, everyone is indifferent between alternative schooling levels, because (discounted) life-time earnings are the same, but there is an earnings premium to each additional year of schooling at every post-schooling age. By equating lifetime earnings for individuals with different levels of schooling, Mincer (1958) derives a log earnings equation that is linear in years of schooling. Mincer (1974) augments the schooling model with a model of on the job investment that leads to a log earnings function that is linear in schooling and quadratic in work experience. That equation has come to be known as the Mincer earnings function, which has had widespread application in empirical work.

---

[139]Heckman and Honore (1990) derive additional implications for earnings distributions and provide identification results in the case of nonnormal distributions.

Mincer's schooling model is silent about which individuals invest in schooling. Becker (1967), in his Woytinsky lecture, specifies a human capital production function in which the marginal return from investing in human capital declines with investment due to an individual's limited capacity. The marginal cost of investing in human capital, the cost of financing additional human capital investment, depends on access to funding (parental and governmental subsidies and loans to education) and is increasing in the level of investment as cheaper sources of capital are used first. The equilibrium level of human capital investment equates the marginal return to marginal cost (at the intersection of the downward sloping demand curve and the upward sloping supply curve). Becker's model implies that the level of human capital investment will in general differ across people, because of differences they face in either supply or demand conditions. For example, a higher level of innate ability implies a higher demand curve, because higher ability is assumed to make human capital investments more productive.

Rosen (1977) translates this framework into a schooling choice model. The log of earnings is assumed to be a function of schooling and ability. Following Becker, schooling (time spent investing in human capital) increases the stock of human capital and thus earnings, but at a decreasing rate. The marginal return is the derivative of the log earnings function with respect to schooling. The marginal cost of schooling, the interest rate at which an individual can borrow depends on family background. The optimal level of schooling is found by equating the marginal return to the marginal cost.

These previous authors model the human capital investment decision as a one-time decision. Ben-Porath (1967) extends the optimal human capital investment decision to a life cycle setting. The Ben-Porath (1967) model assumes that individuals choose a human capital investment profile to maximize discounted lifetime earnings. Human capital is produced at any age through the application of time (a fraction of an individual's human capital stock) and purchased inputs, conditional on an individual's ability and existing stock of human capital. The fraction of the human capital stock not used to produce additional human capital is used to produce earnings. Similar to Roy (1951), an individual's earnings at any age is the product of a market determined price of a unit of human capital and the individual's stock of human capital not used in investment at that age. Schooling, in this framework, is

viewed as a period of full-time investment (no earnings) and on-the-job training as a time of partial investment. Given a finite lifetime, the optimal human capital investment profile, the fraction of time spent investing, declines with age. Thus, any period of full-time investment, that is schooling, would come first.

Willis and Rosen (1979) empirically implement a model of schooling choice that combines the essential features of this early theoretical literature.[140] The paper develops a two sector model where individuals decide whether or not to attend college, basing their decision on expected lifetime earnings with and without college, on financing capacities that differ by family background and on nonpecuniary benefits of education. The model incorporates two unobservable abilities, associated with high school and college level skills. Willis and Rosen (1979) find that the decision to attend college is strongly influenced by expected lifetime earnings gains and that family background is an important determinant of college-going decisions. In addition, they find comparative advantage to be an important feature of the labor market; that is, high school graduates have better prospects as a high school graduate than would an average college graduate and college graduates having better prospects as a college graduate than would an average high school graduate.

2. DCDP models

The DCDP literature extends this earlier work on schooling and occupational choice to a dynamic setting, in which individuals face a sequential decision problem with uncertainty. It incorporates features from the earlier literature, allowing for worker heterogeneity, multiple skill types, latent skills, self-selection and comparative advantage. The literature can be broadly categorized into partial equilibrium approaches, which take skill prices as given, and market equilibrium approaches, where there is an explicit link between the prices paid to skill in the economy and aggregate skill quantities. Here, we first describe partial equilibrium models of schooling and occupational choices and then the more limited set of market

---

[140]Much of the empirical literature prior to Willis and Rosen (1979) was concerned with estimating rates of return to schooling. A considerable amount of effort was (and still is) devoted to accounting for bias in the schooling coefficient due to omitted ability in a Mincer-style earnings function. See Card (2001) and Wolpin (2003) for an assessment of that literature.

equilibrium models. Subsequently, we consider DCDP models that have been developed for particular contexts, for example, to analyze the decision about college major or the decision to enter (and exit) the teaching profession.

Partial equilibrium models of schooling and occupational choice: Gotz and McCall (1984), one of the pioneering papers in the DCDP literature, as noted previously, develops an occupational choice model for the purpose of studying the retention decision of Air Force officers, in particular, how retention responds to compensation policy and to the retirement system. In the model, officers make a binary choice at each age about whether to stay or leave the Air Force so as to maximize the expected present value of pecuniary and nonpecuniary returns. There is a single taste shock that is realized each period and that affects the value of the military option. An officer who leaves joins the civilian labor force and earns a civilian wage. In addition to considering compensation and pension benefits, the model also explicitly accounts for the effects of the chance of promotion on the expected value of staying in the military. The probabilities of promotion and military and civilian earnings are treated as exogenous.[141] The model also allows for persistent differences among individuals in their preference for military service (permanent unobserved heterogeneity).

The model parameters are estimated by maximum likelihood using data on officer employment histories from 1973-1977 as well as data from the Current Population Survey used to construct estimates of civilian earnings. As a way of validating the model, the estimated model is used to forecast retention rates for data not used in estimation, which shows that the estimated model produces good out-of-sample forecasts. The fit of the dynamic retention model is also compared to that of two competing models, one that does not allow for unobserved permanent preference heterogeneity and a lifetime model without per-period shocks where individuals know with certainty the year they will leave the military. The dynamic model that allows both for permanent unobserved heterogeneity and for per-period shocks provides the best fit to the data. The dynamic model is then used to assess the effects of a number of policy interventions of interest to the Military, including (i) an increase in pay

---

[141]Specifically, pay grades and promotions are assumed to be probabilistic functions of observable state variables.

and allowances, (ii) the introduction of a bonus based on years of service completed, (iii) a decline in the value of the military retirement annuity, (iv) an increase in flight pay and (v) indexing pay to the CPI.[142]

Miller (1984), another pioneering paper in the DCDP literature, develops and estimates a matching model of occupational choice. The model assumes that the pay-off to a particular job within an occupation depends on a match-specific component and a random component. Individuals do not know the match-specific component prior to starting the job, but they observe their output. Beliefs about the quality of the match change with experience on the job through a Bayesian updating procedure. Jobs for which the expected return stream are identical are defined as being the same occupation. The model has implications for which jobs should be sampled first and for how long. For example, the notion of equalizing differences would imply that jobs with high informational content pay less on average in equilibrium and attract relatively inexperienced workers who quickly discover their personal match quality and leave in the event of a bad match. These types of jobs would include a large number of inexperienced workers in the process of learning about their match plus a small number of experienced, permanent workers. Jobs with lower informational benefits should have less turnover, pay more for new entrants and have a less variable wage distribution.

The dynamics in the model arise due to the learning process as the value of remaining in a job provides not only an immediate payoff but also information about the future payoff in that job. The parameters of Miller's (1984) job-matching model are estimated by maximum likelihood using data on job tenure and job changes, where the discrete time hazard of remaining on a given job or switching to a new job are derived from solving the dynamic programming model.[143] The hazard function is assumed to also depend on initial observed demographic characteristics of the individual. To capture unobservable heterogeneity, the hazard model depends on two unobserved states, following the approach of Heckman and Singer (1984). The Coleman-Rossi data set, which surveyed a sample of men about their

---

[142] A related study by Daula and Moffitt (1995) uses a similar DCDP model to analyze the effect of financial incentives on Army infantry reenlistments.

[143] The model is solved using the Gittens index, a simplification in the solution of the dynamic programming problem that arises because tenure in one job does not affect the rate of learning in other jobs.

entire work history, education and family background, is used to estimate the model. The empirical evidence supports the prediction that young inexperienced individuals receive low wages in part because they seek out different kinds of occupations, those with greater informational content, than do older individuals.

The preceding papers focused only on the occupational choice decision. The first DCDP model to combine schooling, working and occupational choices in a single framework is Keane and Wolpin (1997). To illustrate concretely the specification of a DCDP model of human capital accumulation, we next describe Keane and Wolpin's (1997) model's structure in detail. As further discussed below, a number of papers in the recent DCDP literature extend the Keane and Wolpin (1997) modeling framework to incorporate additional features.

In the baseline model presented in Keane and Wolpin (1997), individuals make repeated choices over time, starting at age 16 and ending at age 65, about whether to participate in one of five different sectors of the economy: (i) work in a white-collar occupation, (ii) work in a blue-collar occupation, (iii) work in the military, (iv) attend school, or (v) engage in home production. There is a finite horizon during which individuals accumulate schooling and occupation-specific experience that affect future wage earning opportunities.

Denote the five choice alternatives in each time period by m where $m \in \{1, 2, 3, 4, 5\}$. The first three alternatives $\{1, 2, 3\}$ are the work alternatives, the 4th is the schooling alternative and the last is the home alternative. Let $d_{ma} = 1$ if alternative m is chosen in time period a. R_{ma} represents the reward (contemporaneous utility) from choosing alternative m, which captures all benefits and costs associated with that alternative.

The reward in a work sector is the wage, which is the product of the price paid per unit skill times the amount of skill accumulated in that occupational sector. Let r_m denote the rental price paid to skill in occupational sector m and e_{ma} the occupation-specific skill units.

$$R_{ma} = w_{ma} = r_m e_{ma}, \ m \in \{1, 2, 3\}. \tag{177}$$

The technology for skill production depends on the number of years of schooling, g_a, and on occupation-specific work experience, x_{ma}. The production function takes the form:

$$e_{ma} = \exp[e_{m,16} + e_{m_1} g_a + e_{m_2} x_{ma} - e_{m3} x_{ma}^2 + \varepsilon_{ma}], \tag{178}$$

122

where $e_{m,16}$ represents the endowment of skill at age 16. The log wage equation is:

$$\ln w_{ma} = \ln r_m + e_{m,16} + e_{m_1} g_a + e_{m2} x_{ma} - e_{m3} x_{ma}^2 + \varepsilon_{ma} \tag{179}$$

The wage equation has the Mincer form of being linear in years of education and quadratic in experience but has the Ben-Porath (1967) and Griliches (1976) pricing equation interpretation.

If a person goes to school, the per period reward is:

$$R_{4a} = e_{4,16} + \varepsilon_{4a} - t_{c1} 1(attend\,college) - t_{c2} 1(attend\,graduate\,school), \tag{180}$$

where t_{c1} and t_{c2} are tuition costs, $e_{4,16}$, is endowed skill at age 16 and ε_{4a} is a random shock component. The home alternative has the associated nonpecuniary reward:

$$R_{5a} = e_{5,16} + \varepsilon_{5a}, \tag{181}$$

where $e_{5,16}$ is the skill endowment and ε_{5a} the random shock component.

The initial conditions in the model are the highest grade completed at age 16 $(g(16))$ along with the unobserved skill endowments in the different sectors. It is assumed that accumulated experience is zero for all alternatives in the first period. The shock components are assumed to be joint normally distributed and serially independent, conditional on the unobserved endowments:

$$\begin{pmatrix} \varepsilon_{1a} \\ \varepsilon_{2a} \\ \varepsilon_{3a} \\ \varepsilon_{4a} \\ \varepsilon_{5a} \end{pmatrix} \sim N(0, \Omega). \tag{182}$$

The state vector at any a is described by

$$\Omega_a = \{e_{16}, g_a, x_a, \varepsilon_a\}, \tag{183}$$

where

$$e_{16} = [e_{1,16}, e_{2,16}, e_{3,16}, e_{4,16}, e_{5,16}], \tag{184}$$

is the vector of age-16 endowments,

$$x_a = [x_{1a}, x_{2a}, x_{3a}, g_a]. \tag{185}$$

123

is vector of work experience and schooling accumulated in the different sectors and

$$\varepsilon_a = [\varepsilon_{1a}, \varepsilon_{2a}, \varepsilon_{3a}, \varepsilon_{4a}, \varepsilon_{5a}]. \tag{186}$$

is the vector of shocks.

The value function at age a is the maximized value of the expected remaining lifetime utility, taken over all possible sequences of future choices, with respect to the choice at a,

$$V_a(\Omega_a) = \max_{\{d_m(a)\}} E \left[\sum_{t=a}^{A} \delta^{\tau-a} \sum_{m=1}^{5} R_{ma} d_{ma} | \Omega_a \right] \tag{187}$$

The problem can be written recursively in Bellman equation form. For $a < A$, the alternative specific value function is

$$V_{ma}(\Omega_a) = R_{ma}(\Omega_a) + \delta E \left[V_{a+1}(\Omega_{a+1}) | \Omega_a, d_{ma} = 1 \right], \tag{188}$$

where the expectation is taken over the random shock components. In the last time period, A,

$$V_{mA}(\Omega_A) = R_{mA}(\Omega_A). \tag{189}$$

The value function is the maximum over the alternative specific value functions:

$$V_a(\Omega_a) = \max_{m \in M} V_{ma}(\Omega_a) \tag{190}$$

The state variables that evolve over time are the accumulated sector-specific experience and the years of completed schooling:

$$x_{m,a+1} = x_{ma} + d_{ma} \quad m = 1, 2, 3 \tag{191}$$

$$g_{a+1} = g_a + d_{4a} \quad g(a) \leq \bar{g} \tag{192}$$

The observed data are the sector choices that people make and their observed wages (for the sectors with pecuniary rewards) starting from age 16 and ending at age $\bar{a}$ (at most age 27 in the data):

$$[d_{ma}, w_{ma} d_{ma} : m \in \{1, 2, 3\}] \tag{193}$$

$$[d_{ma} : m \in \{4, 5\}]$$

It is assumed that individuals observe contemporaneous shocks ε_a, but that the researcher does not. The observed state space (exclusive of the shocks) is

$$\Omega_a^- = [e_{16}, g_a, x_a]. \tag{194}$$

The likelihood is

$$\Pr[c_{16}, ..., c_a | g_{16}, e_{16}] = \Pi_{a=16}^{\bar{a}} \Pr[c_a | \Omega_a^-], \tag{195}$$

where c_a denotes the vector of choices and wages at age a. The estimation proceeds by: (i) choosing an initial set of parameters, (ii) solving the dynamic programming problem numerically (by approximating the $E\max_a$ functions as previously described), (iii) computing the likelihood, and (iv) iterating to maximize the likelihood until convergence.

The baseline model that Keane and Wolpin (1997) estimate also includes unobservable heterogeneity. Specifically, there are assumed to be 4 types of individuals with heterogeneous age 16 endowments, denoted by

$$e_{k,16} = \{e_{mk,16} : m = 1..5, \ k = 1..4\} \tag{196}$$

The type of the individual is assumed to be known to individuals but unknown to the researcher. Unobservable heterogeneity introduces the potential for comparative advantage into the model in that some individuals persistently get higher rewards in certain sectors, but perhaps not others. Unobserved permanent endowment differences are necessary to fit the high degree of persistence in choices observed in the data.

In the model, the only observable initial conditions that varies is schooling attained at age 16, g_{16}.[144] If the shocks were serially correlated, then it would be problematic to condition the analysis on g_{16}, because g_{16} likely reflects prior schooling decisions that would be affected by earlier shocks. If the shocks are iid, however, then conditioning the analysis on g_{16} is not problematic. The maintained assumption is that the initial condition g_{16} is exogenous with respect to the shocks conditional on the unobserved type. Accounting for unobservable heterogeneity and for initial conditions, the likelihood is:

$$\Pi_{n=1}^N \Sigma_{k=1}^L \pi_{k|g_{n,16}} \Pr[c_{na} | g_{n,16}, type = k] \tag{197}$$

---

[144]Work experience, $x_{16} = 0$.

The type probability is estimated as a function of the initial schooling.

Estimation of the model's parameters is based on eleven years of data on young white males from the NLSY79. The analyses subsample consists of 1373 observations on white males who were age 16 or less as of Oct. 1, 1977 and who are followed through 1988. Each time period in the model corresponds to one year in the data. Wages are measured as full-time equivalent wages, estimated as average weekly wages times 50. Parameter estimates are obtained by simulated maximum likelihood, as previously described.

Keane and Wolpin (1997) evaluate the goodness-of-fit of their baseline model (described previously) and of a few alternative model specifications that differ in their degree of parsimony to learn which features of the model are important to achieving a good fit. Their preferred model augments the baseline model to incorporate skill depreciation during periods of nonwork, job-finding costs, school reentry costs, and nonpecuniary components of work sector alternatives. As a way of validating the model, Keane and Wolpin (1997) also evaluate the fit of the model out-of-sample by predicting the choices of younger birth cohorts (using CPS data) that were not used in estimating the model.

A consistent empirical finding (Willis and Rosen (1979) and Heckman and Sedlacek (1985)) is that comparative advantage plays an important allocative role in the labor market. Workers self-select into occupations and into sectors based on their relative productivities. Keane and Wolpin (1997) find that comparative advantages determined by age 16 are found to lead to large differences in school attainment and later labor market outcomes. Indeed, most of the variation in lifetime utility comes from inequality in skill and preference endowments at age 16, pointing to the importance of early influences in explaining lifetime inequality.

The estimated model is used to predict the effects of a \$2000 (1987 dollars) college tuition subsidy on the college going rate. Under the preferred model specification, the subsidy increases the high school graduation rate by 3.5 percentage points and the college graduation rate by 8.5 percentage points. However, the main beneficiaries of the subsidy, in terms of lifetime utility, are individuals who would have gone to college without the subsidy.

A follow-up paper by Keane and Wolpin (2000) uses a similar framework to analyze the sources of black/white differentials in schooling attainment and earnings and to assess

the impact of policies intended to close the racial gaps. Race enters the model in a number of ways, as a determinant of preference parameters, unobserved type probabilities, and wages. The paper finds that differences in initial age 16 skill endowments are the primary explanation for low schooling attainments of blacks relative to whites. This finding has important implications for policy. Keane and Wolpin (2000) implement a scheme to equalize the schooling distributions of black and white males through the combined use of a high school graduation bonus and a college graduation bonus. Although this policy, by design, closes the racial schooling gap, it has only a very small effect on the racial earnings gap due to differences in skill endowments at age 16.

An area of research that has received much attention in the non-structural literature focuses on the effect of credit market constraints on college enrollment. The finding in that literature that tuition effects are inversely related to parental income has often been interpreted as evidence for the existence of borrowing constraints that have adverse consequences for college attendance (see, e.g., Kane (1999) p. 63). A paper by Keane and Wolpin (2001) studies how borrowing constraints and parental transfers affect educational attainment by estimating a DCDP model of schooling, work and savings decisions of young men, using data from the NLSY79 cohort. The model allows for parents to provide transfers to youths, which the youths take as given and which vary depending on whether the youth chooses to go to college. Like the previous papers, the model incorporates unobserved heterogeneity (endowments at age 16). In the model, schooling and work are not mutually exclusive choices and youths can work full or part time while still attending school (full or part time). Youths may borrow up to a limit. Keane and Wolpin (2001) find that borrowing constraints are tight (financing college tuition through uncollateralized borrowing is not feasible in the model). In addition, consistent with the pattern found in the non-structural literature, Keane and Wolpin report that a tuition increase generates a pattern of larger percentage declines in enrollment for youth whose parents have lower SES.

On the surface, it would appear that the inference drawn in the non-structural literature, that borrowing constraints exist and limit college attendance of youths from less affluent families, is validated by the congruence of these two findings. However, when Keane and Wolpin simulate the impact of relaxing the borrowing constraint, by allowing youths to

borrow the full tuition cost, they find that there is only a negligible increase in college attendance. However, allowing college attendees to borrow up to the full tuition amount leads to a reduction in their propensity to work while attending school and to an increase in their consumption. They therefore conclude that college attendance is not limited to any great extent by borrowing constraints, but rather primarily by age-16 endowments of pre-market skills and/or preferences.[145]

The finding that borrowing constraints are tight and yet relaxing them does not lead to increased college attendance has been controversial. However, it is consistent with earlier research by Cameron and Heckman (1998, 1999) that estimates a sequential model of school attendance decisions.[146] That research finds a strong positive correlation between family income and college attendance, conditional on high school graduation, even after controlling for effects of dynamic selection on unobservables. After controlling for AFQT test score (interpreted as a proxy for the individual's endowment at age 16), however, liquidity constraints no longer play a strong role in college attendance decisions.

In the previously described papers, log wages are specified as a linear function of the number of years of schooling. Belzil and Hansen (2002) estimate a DCDP model of schooling decisions with a focus on allowing the returns to different levels of schooling to vary. In particular, they model the wage equation as a spline in years of schooling with eight knots. Their model assumes that individuals make sequential decisions as to whether to attend

---

[145] Keane and Wolpin (1999b) also find that, on average,.youths receive a transfer from their parents in excess of what is received when not attending college, sufficient to fully subsidize college tuition costs. The subsidy ranges from about one-half of the tuition cost for youths whose parents are the least educated (neither a high school graduate) to almost twice the tuition cost for youths whose parents are the most educated (at least one parent a college graduate). It might appear that it is because of the largesse of parents that relaxing borrowing constraints has only a minimal impact on college attendance. However, simulating the impact of relaxing the borrowing constraint in a regime where parents are assumed to provide no additional transfers to children who attend college leads to the same result. Transfers do, however, have a nonnegligible effect on school attendance. If transfers are equalized across children, high SES children go to school less, but low SES children do not increase their attendance by much.

[146] The approach taken by Cameron and Heckman (1999) can be interpreted as estimating the approximate decision rules from a DCDP model.

school for up to 22 years, after which they enter the labor market.[147] While in school, they receive parental transfers according to a parental transfer function that depends on accumulated years of schooling. After entering the labor market, individuals are employed with some probability and, if employed, receive a wage rate. Both the wage rate and the probability of employment depend on their schooling attainment and labor force experience.[148] The model has three source of uncertainty: a schooling preference shock, a wage shock and an employment shock. It also includes six unobserved types to capture unobservable heterogeneity in schooling ability and in market ability, and one of the goals of the paper is to recover the correlation between unobserved schooling and market ability.

The model is estimated by simulated maximum likelihood on a sample of white males from the NSLY79. The estimated parameters indicate that log wages are convex in years of schooling, with statistical tests rejecting the hypothesis of linearity. The log wage equation has estimated returns to schooling that are very low (1%) until 11th grade, increase to 3.7% in grade 12, and exceed 10% between grades 14 and 16. The estimated returns to schooling are substantially lower than corresponding ols estimates. For a linear in schooling specification, one obtains an ols estimate of 10%, in comparison with the structural model estimates of 2% on average up to grade 12 and 7% after grade 12.[149] They also find that there is a strong positive correlation (0.28) between market ability and realized schooling, which would imply that estimated returns to schooling from wage regressions that do not control for the endogeneity of schooling will tend to upwardly biased.

Sullivan (2007) develops a DCDP model that combines a model of labor force dynamics (as in Wolpin 1992 and Rendon (2006)) with a human capital model of schooling and occupational choice (as in Keane and Wolpin (1997)). The previous literature considered

---

[147]All individuals are assumed to complete at least six years.

[148]In the model, there is also a probability of experiencing a so-called interruption, which is a decision period when no decision is made and the stock of accumulated human capital remains fixed, intended to capture an event such as illness of academic failure that lasts one period.

[149]See Belzil and Hansen (2002) Table VIII. The previously described Keane and Wolpin (1997) model did not allow for nonlinearities in returns to education but did allow the return to differ by white and blue collar occupations and found a much higher return in the white collar occupupation, which could be viewed as consistent with Belzil and Hansen's (2002) finding of high returns at higher education levels.

job search as a separate phenomenon from schooling and occupational choice, though the choices are clearly related. In Sullivan's (2007) model, workers decide in each period whether to attend school and/or work in one of five occupations or neither work nor attend school. An individual who has not graduated from high school may also decide to earn a GED. An employed individual may stay at the current job or switch jobs either within the same occupation or with a change in occupation. Human capital accumulated through work experience is both firm- and occupation-specific. Individuals have heterogeneous skill endowments and preferences for employment in different occupations. Wage offers include a match-specific component, reflecting worker-firm permanent match productivity, and an iid time varying shock. Search arises because of variation in worker-firm match productivity together with mobility costs. Model parameters are estimated by simulated maximum likelihood using data from the NLSY79.

The model estimates are used to perform a number of counterfactuals. Sullivan's (2007) analysis finds that occupational and job mobility are critical determinants of life cycle wage growth, quantitatively more important than the accumulation of occupation-specific human capital. As in previous research, the results also indicate the importance of comparative advantage in understanding schooling and occupational choices. Sullivan also finds that unobservable hetergeneity plays a relatively smaller, though still substantial, role in explaining labor market outcomes than has been found, for example, by Keane and Wolpin (1997). He attributes 56% of the variation in lifetime utility to permanent heterogeneity, which compares to 90% found in Keane and Wolpin (1997).

General equilibrium models of schooling and occupational choice: Most of the literature on modeling occupational and schooling choices is partial equilibrium, taking skill prices as given. However, a few papers in the literature estimate general equilibrium models in which skill prices respond to changes in aggregate market demand and supply for skills.

The earliest paper to estimate a multi-sector general equilibrium model is Heckman and Sedlacek (1985). The paper is an extension of Roy (1951). Although static, and thus not a DCDP model, the paper serves as a link to the later general equilibrium models that fall within the DCDP paradigm. Specifically, Heckman and Sedlacek (1985) estimate a model

130

of individuals' decisions among three sectors: work in the manufacturing sector, work in the nonmanufacturing sector or not work. In addition to specifying the micro-level supply-side sector choice model, the paper estimates an aggregate demand function for skill. The micro supply-side model and the aggregate demand models are used jointly to simulate the effects of price changes on employment levels and wages, such as an increase in the price of energy that predominantly affects labor demand in manufacturing.

Establishing the link between aggregate skill quantities and skill prices can be conceptually important for analyzing the effects of policy interventions. To illustrate, consider, for example, the impact of a college tuition subsidy on the fraction of people going to college. A tuition subsidy must act as a positive inducement to college attendance. In a general equilibrium framework, a college tuition subsidy that induces more people to go to college would also decrease the price paid to college skill given the increase in the aggregate quantity of college educated labor. For this reason, we would expect the predicted general equilibrium effect of a college tuition subsidy on college-going to be smaller than the predicted partial equilibrium effect. The quantitative significance of the supply effect on skill prices is an empirical question.

The papers described below develop and estimate general equilibrium models incorporating schooling and employment choices. The goals of these papers are to understand historical wage and employment patterns for workers of different skill levels and to analyze the effects of skill formation policies, such as tuition subsidies.

Wage inequality has increased at least since the 1980's, with low skill workers experiencing both absolute and relative declines in real wages as the economic returns to skill acquisition have risen. Heckman, Lochner and Taber (1998) (HLT) present the first general equilibrium model of schooling and job training choices, which they use to explore alternative explanations for observed wage patterns and to simulate the effects of college tuition subsidies. In the HLT model, individuals make decisions about whether to go to college, about post-school on-the-job training (human capital investments a la Ben-Porath) and about life-cycle savings. The model assumes that individuals decide whether or not to go to college and on their optimal life cycle consumption and human capital investment paths, assuming they work

each period until the age of retirement.[150] There are no credit constraints. The market wage for each skill type is the product of the skill rental price and the amount of accumulated skill. Individuals are heterogeneous in terms of initial skill endowments, captured by the observed AFQT test score.

The model is solved for overlapping generations of agents and estimated using both aggregate CPS data (from 1963 to 1993) and longitudinal data from the NLSY79. Cohorts make different choices because they face different (known) skill prices over their lifetimes. The model assumes a one-to-one correspondence between schooling groups (high school and college) and skill types, that is, that different skill types cum schooling classes are imperfectly substitutable. However, age groups within a given schooling group (high school or college) are perfect substitutes. Skill prices are determined in equilibrium. Equilibrium skill prices induce aggregate skill supplies that equate marginal revenue skill products to skill prices.

HLT calculate the partial equilibrium and general equilibrium impact of a 100 dollar increase in tuition on college enrollment. They find that the partial equilibrium response is a decline in enrollment of 1.6 percent. However, when they allow for skill prices to adjust to the reduction in college skill, that is for the increase in the relative price of college skill, they find that the decline in enrollment is less than 0.2 percent. Thus, the adjustment in the relative price of college to high school skill almost completely offsets the disincentive to acquire schooling. Presumably, a tuition subsidy of a similar magnitude can be expected to lead to only a negligible increase in college enrollment due to the fall in the relative college skill price.

Lee (2005) estimates an alternative formulation of a general equilibrium schooling and occupational choice model. The specification of the individual's problem parallels that of Keane and Wolpin (1997). Specifically, in each period an individual decides whether to attend school, work in one of two occupations, blue collar or white collar jobs, or do neither. Individuals are heterogeneous in skill and preference endowments and are subject to idiosyncratic time-varying shocks. A critical difference between the model of Lee (2005) and that of HLT (1998) is that in Lee's model, an individual's skill type is not equated with their

---

[150]The model is life cycle rather than dynamic in the sense that new information is not revealed to the agent in each decision period.

schooling. Schooling augments both white and blue-collar skill, though differentially, and it is the aggregate levels of the occupation-specific skills that enter as inputs into the aggregate production function. In Lee's (2005) model, occupations are not perfectly substitutable, but education types are perfectly substitutable within occupation and age groups. This difference between the HLT (1998) and Lee (2005) models has important consequences for the relationship between partial and general equilibrium effects of tuition policies.

Lee (2005) estimates the model using simulated method of moments applied to CPS data on schooling, occupational choice, employment and cohort size, under an assumption that individuals have perfect foresight about future skill prices. The estimated model is then used to investigate how cohort size affects skill prices and wages and also to evaluate the effects of a college tuition subsidy. As in HLT, Lee evaluates the partial and general equilibrium impacts of a 100 dollar increase in tuition on college enrollment. The partial equilibrium effect ranges from 1.2 to 1.9 depending on age and gender, similar in magnitude to HLT. However, the general equilibrium effect is only found to be about 10 percent lower, in sharp contrast to the result in HLT. The reason, supported by simulations performed by Lee, is that workers can respond to changes in the relative price of blue- and white-collar skill by switching sectors. A tuition increase that reduces college enrollment, and thus increases the relative white collar skill price, induces some blue-collar workers to switch to white-collar jobs, mitigating the rise in the white collar skill price.

A recent paper by Lee and Wolpin (2009) develops and estimates a general equilibrium model to explain the evolution of wages and employment over the last 30 years, including gender differentials in employment and earnings, which were not considered in the previous two studies that focused only on males.[151] Specifically, the study aims to account for changes in wage inequality (both overall and within demographic groups), increases in relative wages and employment of women, and a shift that has occurred over time in employment from the goods to the service producing sector. There is an extremely large, mostly nonstructural, literature that considers each of these major labor market changes as separate phenomena.[152]

---

[151] An earlier paper by Lee and Wolpin (2006) develops a similar model with a focus on examining the relative importance of labor demand and supply factors in explaining the expansion of service sector employment

[152] The general rise in inequality and the college premium have often been linked (for example, Murphy and Welch (1992)), but not together with the rise in femal-male wages and the growth of the service sector.

Lee and Wolpin (2009) develop a comprehensive framework which includes many of the factors considered to be potential explanations for these major labor market changes.[153]

The model estimated in the paper has two production sectors, corresponding to goods and services. Aggregate production depends on three skill types (white-, pink- and blue-collar) and on capital. There are time-varying neutral and non-neutral technological changes as well as combined aggregate productivity and relative product price shocks. The goods-to-service product price and the price of capital evolve exogenously.

In the model, men and women age 16-65 can choose to work in any of six sector-occupations (pink collar, white collar or blue collar in either the goods or service sectors), to attend school or to stay home. Each period, individuals receive wage offers from each sector-occupation that depend on schooling attainment and accumulated experience in each sector-occupation. There are also nonpecuniary payoffs and preference shocks to the different options. To capture lower labor force participation rates of women during child-bearing ages, the value of the home choice is assumed to depend on the number of preschool age children in the household. It is also allowed to vary over time to reflect technological improvements that are thought to have occurred in the home sector. In addition, there is a cost of transiting between sector-occupations, which can be interpreted as labor market frictions.[154] The population at any point in time consists of overlapping generations of both sexes. Unobservable heterogeneity is incorporated by including five unobserved types of individuals who differ in sector-specific endowments and in preferences for the home and school options.

Skill prices are equated to marginal revenue products evaluated at aggregate skill amounts. The paper also develops a belief consistent forecast rule for future skill prices, as an approximation of a rational expectations equilibrium. Model parameter estimates are obtained by

The growth in the service sector has also been linked with the rise in female employment (Fuchs (1980)) and Welch (2000) draws a link between the ride in wage inequality among men and the reduction in the gender wage gap.

[153] For a review of the larger literature, see Katz and Autor (1999) The papers they survey include Bound and Johnson (1992), Gottschalk and Moffitt (1994), Juhn, Murphy and Pierce (1993), Katz and Murphy (1992), Krusell et. al. (2000) and Murphy and Welch (1992, 1993). Recent contributions to this literature include Baldwin and Cain (2000), Eckstein and Nagypal (2004), Hornstein et. al. (2004) and Welch (2000).

[154] There is a large labor economics literature on interindustry wage differentials among otherwise observably identical workers. Frictions to switching sectors are sometimes proposed to explain these differentials.

simulated method of moments, matching the model's predicted levels of wages, employment and school enrollment to data from the CPS, BLS and NLSY79.

Lee and Wolpin (2009) use the estimated model to assess the relative contribution of changing technology, preferences and exogenous forcing variables (the goods to service product price, the price of capital, fertility) as explanations of the previously described major labor market changes. This is done by using the model to simulate labor force outcomes under hypothetical scenarios relative to a baseline economy. The key findings from the analysis are that (i) neutral technological change best accounts for service sector employment growth, (ii) skill-biased technological change best explains the rise in the college wage premium and the increase in overall wage inequality, (iii) the combination of neutral and biased technological change account for the declining gender gap and increased female labor force participation, and (iv) changes over time in fertility and in the valuation of the home sector can account for female-male relative wage and employment growth. The study concludes that a competitive general equilibrium model of the labor market provides a comprehensive framework for analyzing the determinants of wage and employment changes over the last 30 years and that both demand and supply side factors are required to account for the major labor market changes.

3. The use of DCDP models in related contexts

The previously discussed studies focused on schooling and occupational choice decisions. We next describe a DCDP literature that develops models to study how marriage decisions interact with labor force decisions, the operation of particular occupational labor markets, various behaviors of adolescent youth and the effects of job training on training program participants.

Marriage and career decisions: Gould (2008) estimates a DCDP model of marriage and career decisions of young men age 16-35 using data from the NLSY79 with the aim of exploring the extent to which schooling and employment choices are influenced by marriage market considerations. Individuals choose among four sector options: schooling, white-collar work, blue-collar work and home. In addition, men face potential marriage opportunities that

135

are conditional on their demographic characteristics and on their current and previous marriage, schooling and employment decisions. Based on the available opportunities, they decide among marriage states. Specifically, there is some probability of receiving a marriage offer (from a woman of given type) and men decide whether to accept the offer. Married men also face an exogenous probability of having their marriage terminated by their wife. The model incorporates four unobserved types of individuals to allow for unobservable heterogeneity.

Model parameters are estimated by simulated maximum likelihood.[155] The estimated model is used to study how young men's career choices would change if there were no marriage market returns to career decisions, that is, by shutting down marriage within the model. Simulation results show that the marriage market significantly affects men's schooling and labor market decisions. Without marriage, men work less, study less, and relatively more often choose blue-collar over white-collar work. Another simulation examines the effects of changing divorce costs on men's choices. A decrease in divorce costs leads men to take fewer measures to guard against a marital break-up; they invest less in education and relatively more often choose blue-collar over white-collar work. Overall, Gould (2008) finds that the private returns to human capital investment include significant returns in the marriage market.

Occupational labor markets: We next describe three studies that use DCDP models to analyze the operation of a particular occupational labor market. Sauer (1998, 2004) studies life-cycle career choices of law school graduates following graduation from the University of Michigan Law School and how these choices are affected by financing options and loan forgiveness programs. Stinebrickner (2001) studies the decisions by certified elementary and secondary teachers to stay in or exit from the teaching sector.

Sauer (1998) estimates a model of a law school graduate's choices among five employment sectors that differ in pecuniary and nonpecuniary returns, in promotion and dismissal probabilities, and in the extent to which human capital is transferable across sectors. The possible employment sectors are nonprofit, elite private law firm, nonelite private law firm,

---

[155] The maximum likelihood approach, developed by Keane and Wolpin (2001) and extended in Keane and Sauer (2002) allows for classifiction error and for missing state variables.

separate business and sole proprietor. Attornies choose among the different sectors taking into account effects of current choices on future job opportunities and wage offers, which depend on endogenously accumulated sector-specific work experience. Within private law firms, lawyers also have the opportunity for promotion from an associate to a partner position. The model includes unobserved types that are assumed to be known to both the worker and to the firm, but not to the researcher. The model is estimated by simulated maximum likelihood.

An interesting aspect of the model is that it generates sector-specific nonmonotonic hazards in duration of employment, as observed in the data, through a mechanism that is different from that of the classical job-matching model (Jovanovic, 1979). In Sauer's (1998) model, the ability of the worker and the quality of the match are known from the beginning, and nonmonotonic hazards arise because of self-selection. In particular, high ability lawyers face higher probabilities of promotion at private law firms and stay at these firms when they get promoted. Low ability lawyers initially also work at private firms even though they have a low probability of getting promoted, because their experience pays off later in the form of higher-paying jobs in other sectors. The self-selection mechanism has implications for effects of policy interventions in the market for lawyers, such as programs that forgive loans if a lawyer enters the nonprofit sector. Simulations using the estimated model indicate that a loan forgiveness program induces low ability types to enter the nonprofit sector earlier but is relatively ineffective in attracting high ability lawyers.

A follow-up paper by Sauer (2004) extends his previous DCDP model to incorporate educational financing decisions. The study's goal is to measure the effects of short-term parental cash transfers and family background on educational borrowing and in-school work decisions, and ultimately on earnings after graduation, and also to better understand effects of policies such as tuition tax credits and loan forgiveness programs on these decisions and on post-graduation outcomes. The model assumes that individuals maximize their expected present value of lifetime utility by making decisions on the level of educational indebtedness, whether to work while in school and the type of post-graduation employment. Total financial resources during law school come from five possible sources: parental cash transfers, initial assets, stochastic unobserved assets, educational debt, and stochastic labor income.

137

Post-graduation job market choices are modeled analogously to Sauer (1998). The model also includes three unobserved types to capture unobservable heterogeneity, where the type probabilities depend on family background variables (that include whether the father was an attorney) and whether the individual has an Ivy League BA.

Model parameters are estimated by simulated maximum likelihood allowing for classification error. The estimated model is used to examine the effects on student's borrowing, work while in school and subsequent employment choices of a loan forgiveness program that grants an annual subsidy equal to an individuals' debt obligation for individuals who take jobs in the non-profit sector within the first 10 years after graduation. Simulations of the model with and without such a program indicate that the program increases borrowing and reduces work while in school. However, the loan forgiveness program has essentially no effect on the choice of first job, with the same types of individuals being most likely to enter the non-profit sector.[156] The main difference is that they enter that sector with more debt. The effect of the loan forgiveness program on participation in the nonprofit sector is similar to that found in Sauer (1998), except that now allowing for individuals to change their borrowing behavior increases substantially the cost of providing the program.

Stinebrickner (2001) develops and estimates a DCDP model to study the decision of certified elementary and secondary teachers to remain in the teaching sector, to exit into the nonteaching sector or to leave the labor force. Certified teachers often leave the teaching sector within two to nine years following certification. This high turnover is of particular concern because certification requirements and wage rigidities in the teacher labor market make it difficult to adapt to fluctuations in teacher demand.

In Stinebrickner's (2001) model, certified teachers receive wage offers in each period in both the teaching and nonteaching sectors. They decide whether to work in the teaching sector, in the nonteaching sector or to not work. The model incorporates marital status and number of children, which are assumed to evolve exogenously. Also, wage offers in both the teaching and nonteaching sectors are allowed to depend on the individual's SAT score, interpreted as a measure of academic ability. Model parameters are estimated by simulated

---

[156]These results have the same flavor as those in Keane and Wolpin (2001) in their simulations of the effect of reducing borrowing constraints.on college enrollment.

maximum likelihood using data on 450 certified teachers from the National Longitudinal Survey of the High School Class of 1972.

A key result from the analysis is that the primary cause of leaving decisions by teachers is not the relative attractiveness of nonteaching occupations but rather the decision not to work, which for women is strongly influenced by changes in marital status and in numbers of children. For teachers with a high SAT score, though, relatively better options in the nonteaching sector, which has a larger earnings premium for skills measured by the SAT, is a factor influencing their decision to leave the teaching sector. Model simulations indicate that teacher labor supply is responsive to changes in teaching wage offers. Increasing the teacher wage by 20 percent increases the proportion of person-years spent in teaching from 0.5 to 0.8, with a greater response among teachers with higher SAT scores.

Schooling-related choices: There have been a few applications of DCDP models to analyze youth behaviors while in school, for example, the decisions by youth to work while in school, drop-out of school, to enroll in college or to major in a particular subject in college. Eckstein and Wolpin (1999) use a DCDP model to study the determinants of school dropout decisions and to analyze whether working while in school is detrimental to school performance. In the model, youths choose among various work-school combination alternatives so as to maximize expected lifetime utility. Youths who attend high school accumulate credits towards graduation and receive grades reflecting their performance. In each period, they also receive random wage offers for either part-time or full-time employment, which they can either accept or reject. The wage offers depend on their skill endowments, educational attainment and previous labor market experience. Working potentially reduces school performance, as measured by course grades, and thus may increase the probability of failing to progress. The model also incorporates unmeasured heterogeneity at the time of entering high school, in preferences, abilities and in the expected value assigned to receiving a high school diploma. The model is estimated by simulated maximum likelihood using white males from the NLSY79.

Determining the impact of work on high school performance has been the subject of a substantial economics, sociology, and psychology literature.[157] Eckstein and Wolpin (1999)

---

[157]See Greenberger and Steinberg (1986).

finds that working while in school reduces academic performance, but the quantitative effect is small. A hypothetical policy that forces youths to stay in high school for five years or until they graduate without working increases the percentage of high school graduates by only 2 percentage points, but increases the average number of years of high school completed by dropouts by one year. As in other studies, initial traits at the time of starting high school are found to be major determinants of dropping out behavior. Youths of the types with lower school ability and/or motivation, a lower expected value of a high school diploma, a higher value placed on leisure time, higher skills in jobs that do not require a high school diploma, and a lower consumption value of attending school tend to drop out of school. The implication is that youth labor policies that do not alter the traits that youths bring to high school will be relatively ineffective in improving school outcomes.

Arcidiacono (2005) uses a DCDP model to study how changing the admission and financial aid rules at colleges affect future earnings of individuals. Specifically, he develops and estimates a behavioral model of decisions about where to submit college applications, which school to attend and what field to study. In the model, individuals make application decisions based on their expectation of the probability of acceptance, the application cost, the expected financial aid conditional on acceptance, and on an expectation of how well they will like a particular college and major combination. Schools make admissions and financial aid decisions; but rather than specifying and structurally estimating the school optimization problem, it is assumed that the school's maximization problem leads to a logit probability of a particular student being admitted to school conditional on the quality of the school and the individual's own ability. School quality is measured by the average math and verbal SAT at the school.[158]

Conditional on the offered financial aid and acceptance set, individuals decide what school to attend and what field to study. They also have the option of not attending school and going directly to the labor market. After college, individuals enter the labor market and their expected utility is equated to the log of the expected present value of lifetime earnings. The model is estimated using a panel data on high school graduates from a single cohort (the

---

[158]Recent work that specifies and estimates equilibrium models of the college market include Epple, Romano and Sieg (2003) and Fu (2010).

National Longitudinal Study of the Class of 1972). Parameters are estimated by simulated maximum likelihood.[159]

The estimated model is used to examine (i) the effects of affirmative action on college-going decisions of African American students and on their labor market outcomes and (ii) the reasons for large earnings and ability differences across college majors. With regard to affirmative action, Arcidiacono (2005) simulates how African American educational choices would change if they faced white admission and financial aid rules. Past research has shown that racial preference in the admissions process is a practice mainly at top tier institutions. Model simulations show that removing racial advantages in financial aid substantially reduces the number of African Americans who attend college and that removing advantages in admission reduces the number attending top-tier schools. However, even though such policies affect the college choice decision, they do not do much to alter lifetime earnings, which is in large part determined by initial endowments, in line with Keane and Wolpin's (1997) earlier finding.

Second, Arcidiacono (2005) uses the model to examine the reasons for large earnings and ability differences across college majors, in particular the high earnings premiums for natural science and business majors. Arcidiacono (2005) finds that monetary premia for certain majors cannot explain ability sorting across majors. Instead, almost all of the sorting occurs because of differing preferences for majors (and the jobs associated with those majors) by initial abilities. Differences in math ability is shown to be an especially important factor explaining both labor market returns and sorting across majors.

Job training: Cohen-Goldner and Eckstein (2009) uses a DCDP model to study the impact of a job training program on labor mobility and human capital accumulation. Their data consist of a short panel of observations on 419 prime age male immigrants in Israel who came from the former Soviet Union. Many of these immigrants were highly skilled upon arrival to Israel, but some of their skills were not directly transferable to the Israeli labor market. A typical pattern in the data is that immigrants start out as unemployed, move to blue collar

---

[159]Estimation is based on the EM algorithm developed in Dempster, Laird and Rubin (1977) and adapted to DCDP models by Arcidiacono and Jones (2003).

jobs and then gradually move into white collar jobs. The government offers these immigrants a language course and job training courses to facilitate their employment transition, with a requirement that they pass a test in the Hebrew language to participate in training. One of the goals of Cohen-Goldner and Eckstein (2009) is to study the effects of these local training courses on labor market outcomes.

To this end, the paper estimates a DCDP model in which immigrants can be in one of the following states: employed in a blue-collar occupation, employed in a white-collar occupation, attending a training course in a blue-collar occupation, attending a training course in a white-collar occupation, or unemployed. An immigrant's language ability (which is self-reported) is assumed to evolve exogenously. In each period, individuals have some probability of receiving a white- or blue-collar job offer and of receiving a white-collar training offer. Blue-collar training is always an option for those with sufficient knowledge of Hebrew. Wage offers depend on the immigrants' accumulated human capital, which in turn depends on accumulated experience and training, language fluency and imported skills. The model includes four unobserved types to capture unobservable heterogeneity. It is estimated by simulated maximum likelihood.

Using the model, Cohen-Goldner and Eckstein (2009) estimate the returns from job training, controlling through the decision model for selectivity into training. They find that participating in white-collar training increases mean accepted wages by 6% and blue-collar training by 9.8%. Participating in white-collar training also doubles the white-collar job offer probability. The net present value of government sponsored training to the immigrants is estimated to be 2.8-3.7%.

Models of Early Childhood Investments: As earlier described, much of the work on understanding the sources of inequality in lifetime earnings and utility among individuals emphasizes the role of pre-market factors. A small literature in economics seeks to model in a dynamic setting how parental investments affect human capital formation during childhood and adolescent years and how parental investment levels are chosen. Bernal (2008) develops and estimates a DCDP model of employment and child care decisions of women, using the NLSY-Children dataset. She finds that mother's employment status and child care choices

142

affect children's cognitive achievement outcomes, as measured by scores on the PPVT and PIAT (math and reading). Having a mother who works full-time and uses child care reduces test scores by about 2%. She uses the model to explore the effects of policy interventions that include child care subsidies and maternity leave entitlements, which she finds, on average, have adverse affects on pre-school aged children's cognitive outcomes. For example, a 35% child care subsidy, which increases the labor supply of mothers, reduces test scores by 0.23%-1.8%, depending on the age of the child.

Recent work by Cunha, Heckman, and Schenach (2009) develops new methods for estimating the so-called "technology of human capital formation." In particular, they develop and estimate a multistage, dynamic model of the evolution of cognitive and noncognitive skills as determined by parental investments in different periods of a child's life. As in Bernal (2008), they use the NLSY-Children data and measure cognitive skills by math and reading scores. Noncognitive skills would include such factors as motivation, self-efficacy, stubbornness and temperament, for which empirical measures are also available. The paper posits a conceptual framework in which current skills are a function of previous skill levels and intervening parental investments, parental skills, and unobserved components. It is assumed that parents sequentially choose investment levels to maximize their child's expected net present value of lifetime earnings, which implies that investment is endogenous in the estimation of the skill production technology. The paper develops an approach for addressing the endogeneity problem that jointly estimates the skill production function along with the investment policy function that can be viewed as an approximation to the decision rule from the dynamic programming problem.

A major goal of the paper is to recover substitution parameters that govern the relative important of early verses late parental investment for subsequent lifetime achievement. Cunha, Heckman and Schenach (2009) find that investments in the early years are particularly important for the formation of adult cognitive skills and that investments in cognitive skill are much more productive if applied at younger ages. On the other hand, for noncognitive skills, there are not dramatic differences in the productivity of investments across different life-cycle stages.

4. Summary

There is a rather consistent finding that human capital accumulation that has already occurred by the age of college attendance decisions plays a large role in subsequent labor market outcomes. Thus, policies like tuition subsidies, student loans, affirmative action, etc., aimed at affecting college attendance, may come too late in the life-cycle to remediate the main factors that lead to inequality in outcomes.[160] The literature on schooling began with and has always emphasized the college attendance decision (see Walsh, 1935). If the goal is to understand the determinants of inequality and the effects of policies for reducing inequality, then the results to date strongly suggest that the focus should shift to human capital investment decisions at earlier ages.

V. Concluding Remarks - How Credible are DCDP Models?

As we have illustrated, applications of the DCDP approach have addressed challenging and important questions often involving the evaluation of counterfactual scenarios or policies. The ambitiousness of the research agenda that the DCDP approach can accommodate is a major strength. This strength is purchased at a cost. To be able to perform such counterfactual analyses in such a variety of settings, DCDP models must rely on extra-theoretic modeling choices, including functional form and distributional assumptions. It is tempting to dismiss the approach for that reason, although we see no other empirical methodology with which to replace it. All approaches fall short of an assumption-free ideal that does not and is likely never to exist. And, as we have noted, DCDP researchers have taken the need to provide credible validation seriously.

There are two approaches to model validation, stemming from different epistemological perspectives. The first is the view that knowledge is absolute, that is, there exists a "true" decision-theoretic model from which observed data are generated. This leads naturally to a model validation strategy based on testing the validity of the model's behavioral

---

[160] The DCDP schooling models described previously take endowments at college entry ages as given. It is still an open question whether college subsidies would induce an increase in parental investments at younger ages and thus affect the endowments.

implications and/or testing the fit of the model to the data. A model is deemed invalid if it is rejected according to some statistical criterion. Rejected or invalid models are discarded.

The second approach is based on a pragmatic epistemological view, in which it is acknowledged that all models are necessarily simplifications of agents' actual decision-making behavior. Hypothesis testing as a means of model validation or selection is eschewed because, given enough data, all models would be rejected as true models. In this pragmatic view, there is no true decision-theoretic model, only models that perform better or worse in addressing particular questions. Models are chosen that are "best" for some specific purpose; alternative models may be valid for different purposes.

Decision-theoretic models are typically designed and estimated with the goal of predicting the impact on economic agents of changes in the economic environment. Thus, one criterion for model validation/selection that fits within the "pragmatic" view is to examine a model's predictive accuracy, namely, how successful the model is at predicting outcomes of interest within the particular context for which the model was designed. In contrast, in the absolutist view, a model would be considered useful for prediction only if it were not rejected on statistical grounds, even though non-rejection does not necessarily imply predicted effects will be close to actual effects. Nor will non-rejected models necessarily outperform rejected models in terms of their (context-specific) predictive accuracy.

Can one provide convincing evidence about the credibility of these exercises? Put differently, how can DCDP models be validated and choices be made among competing models?

There are a number of possible approaches to model validation/selection.

1. Robustness to assumptions: One method is to check how robust the empirical findings are to alternative assumptions. Although, in principle, such a method would provide evidence on the credibility of any particular set of findings, the number of assumptions in these models, their computational burden, and disagreements among researchers as to the *a priori* importance of particular assumptions, has led practitioners to limit the use of this approach.[161] That is not to say that evidence accumulated from the estimation of models by different researchers, each with different modeling inclinations, is not valuable.[162] Indeed,

---

[161] Of course, robustness by itself cannot be conclusive; all of the models could give similarly biased results.

[162] Such prejudices are revealed by the contrast between the structure of the DCDP model that Todd

contradicting findings could be very revealing.

2. Within-sample model fit: DCDP papers commonly assess model fit to the estimation sample, often, but not always, using formal statistical tests. The problem with basing validation on model fit is that, like non-structural estimation, model building is an inductive as well as deductive exercise. The final specification results from a process in which the model structure is revised as estimation proceeds, by adding parameters and changing functional forms, as deficiencies in model fit are discovered. This process of repeated model pre-testing invalidates the application of standard formal statistical tests. Nevertheless, it is interesting to note that formal tests generally reject DCDP models. Although these models tend to have a lot of parameters, sometimes numbering into the hundreds, given the extensiveness of the data moments that these models attempt to fit, the models are actually parsimonious. In practice, researchers tend to stop well short of fitting the model to idiosyncratic features of the data just to improve model fit.

3. Out-of-sample validation: Out-of-sample validation relies on there being sample data not used in estimation, but that is assumed to come from the same underlying population. This *validation* sample can have a number of sources. One source for the validation sample is based on regime shifts. McFadden (1977), for example, estimated a random utility model (RUM) of travel demand before the introduction of the San Francisco Bay Area Rapid Transit (BART) system, obtained a forecast of the level of patronage that would ensue, and then compared the forecast to actual usage after BART's introduction. McFadden's model validation treats pre-BART observations as the estimation sample and post-BART observations as the validation sample. The validation exercise exploited data that were unavailable for estimation to validate their model.

Some researchers have deliberately held out data to use for validation purposes. Lumsdaine, Stock, and Wise (1992), for example, estimated a model of the retirement behavior of

---

and Wolpin (2006) used to evaluate a conditional cash transfer program in Mexico and the model used by Attanasio, Meghir and Santiago (2005). As another example, there are several applications of DCDP models applied to traditional topics that take a behavioral economics view. As seen, Paserman (2008) studies a job search model. In addition, Fang and Silverman (2005) study a model of women's welfare participation assuming that agents use hyperbolic discounting.

workers in a single firm who were observed before and after the introduction of a temporary one-year pension window. They estimated several models on data before the window was introduced and compared the forecast of the impact of the pension window on retirement based on each estimated model to the actual impact as a means of model validation and selection. Keane and Moffitt (1998) estimated a model of labor supply and welfare program participation using data after federal legislation (OBRA 1981) that significantly changed the program rules. They used the model to predict behavior prior to that policy change. Keane and Wolpin (2007) estimated a model of welfare participation, schooling, labor supply, marriage an fertility on a sample of women from five U.S. States and validated the model based on a forecast of those behaviors on a sixth state. The validation sample was purposely drawn from a state in which welfare benefits were significantly lower than in the estimation sample.

Randomized social experiments have also provided opportunities for model validation and selection. Wise (1985) exploited a housing subsidy experiment to evaluate a model of housing demand. In the experiment, families that met an income eligibility criterion were randomly assigned to control and treatment groups. The latter were offered a rent subsidy. The model was estimated using only control group data and was used to forecast the impact of the program on the treatment group. The forecast was compared to its actual impact. More recently, Todd and Wolpin (2006) used data from a large-scale school subsidy experiment in Mexico, where villages were randomly assigned to control and treatment groups. Using only the control villages, they estimated a behavioral model of parental decisions about child schooling and work, as well as family fertility. The validity of the model was then assessed according to how well it could forecast (predict) the behavior of households in the treatment villages.[163]

As should be clear from this discussion, model validation, and model building more generally, are part art and part science. For this reason, researchers will attach different priors to a model's credibility, different weights to the validation evidence, and may, therefore, come

---

[163]Similarly, Lise, Seitz and Smith (2003) used data from a Canadian experiment designed to move people off of welfare and into work to validate a calibrated search-matching model of labor market behavior. Bajari and Hortacsu (2005) employ a similar validation methodology in the case of a laboratory experiment.

to different conclusions about the plausibility of the results. Presumably, disagreements can be reduced as confirmatory or contradictory evidence is accumulated. Whatever empirical approach to inference is adopted, structural or non-structural, researchers should strive to provide as much validation evidence as the data and methods permit.

References

[1] Aguirregebaria, V. and P. Mira (forthcoming). "Dynamic Discrete Choice Structural Models: A Survey." *Journal of Econometrics.*

[2] Altug S. and R. A. Miller (1990). "Household Choices in Equilibrium". *Econometrica* 58: 543-570.

[3] Altug S. and R. A. Miller (1998). "The effect of work experience on female wages and labor supply". *Review of Economic Studies* 65: 45-85.

[4] Albrecht, J. and B. Axell (1984). "An equilibrium model of search employment". *Journal of Political Economy* 92: 824-840.

[5] Arcidiacono, P. (2003). "Ability sorting and the returns to college major". *Journal of Econometrics* 121: 343-375.

[6] Arcidiacono, P. (2005). "Affirmative action in higher education: How do admission and financial aid rules affect future earnings?" *Econometrica* 73: 1477-1524.

[7] Arcidiacono, P. and Jones, J. B. (2003). "Finite mixture distributions, sequential likelihood, and the EM algorithm". *Econometrica* 71: 933-946.

[8] Arcidiacono, P. and R.A. Miller (2008). "CCP estimation of dynamic discrete choice models with unobserved heterogeneity". mimeo, Duke University.

[9] Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* New York: National Bureau of Economic Research.

[10] Becker, G. S. (1967). "Human capital and the personal distribution of income: An analytical approach". Woytinsky Lecture no. 1. Ann Arbor: University of Michigan, Institute of Public Administration.

[11] Becker, G. S. and B. R. Chiswick (1966). "Education and the distribution of earnings". *Journal of Political Economy* 56(1/2): 358–369.

[12] Bellman, R. (1957). *Dynamic Programming.* Princeton: Princeton University Press.

[13] Bellman, R., R. Kalaba and B. Kootkin (1963). "Polynomial approximation - A new computational technique in dynamic programming allocation processes". *Mathematics of Computation* 1: 155-161.

[14] Belzil, Christian (2007). "The return to schooling in structural dynamic models: a survey". *European Economic Review* 51: 1059-1105.

[15] Belzil, C. and J. Hansen (2002), "Unobserved ability and the return to schooling". *Econometrica* 70: 2075-91.

[16] Ben-Porath, Y. (1967). "The production of human capital and the life-cycle of earnings". *Journal of Political Economy* 75: 352-365.

[17] Bound, J. and G. Johnson (1992). "Changes in the structure of wages in the 1980s: An Evaluation of Alternative Explanations". *American Economic Review* 82: 371-92.

[18] Blau D (1991). "Search for nonwage job characteristics: A test of the reservation wage hypothesis". *Journal of Labor Economics* 9: 186-205.

[19] Bernal, R. (2008). "The effect of maternal employment and child care on children's cognitive development." *International Economic Review* 49: 1173-1209.

[20] Blundell R.W. and I. Walker (1986). "A life cycle consistent empirical model of labor supply using cross section data". *Review of Economic Studies* 53: 539-558.

[21] Blundell R.W., A. Duncan and C. Meghir (1998). "Estimating labour supply responses using tax policy reforms". *Econometrica* 66: 827-861.

[22] Bontemp, C., J-M Robin and G.J. van den Berg (1999). "An empirical equilibrium job search model with search on the job and heterogeneous workers and firms". *International Economic Review* 40: 1039-1074.

[23] Bontemp, C., J-M Robin and G.J. van den Berg (2000). "Equilibrium search with continuous productivity dispersion: Theory and nonparametric estimation". *International Economic Review* 41: 305-358.

[24] Burdett, K. and D. T. Mortensen (1998). "Wage differentials, employer size and unemployment". *International Economic Review* 39: 257-273.

[25] Cahuc, P., F. Postel-Vinay and J-M. Robin (2006). "Wage bargaining with on the job search: Theory and evidence." *Econometrica* 74: 323-364.

[26] Card, D. (2001). "Estimating the Returns to Schooling: Progress on Some Persistent Econometric Problems". *Econometrica* 69: 1127-1160.

[27] Ching, A, Imai, S., Ishihara, M. and N. Jain (2010). "A guide to Bayesian estimation of discrete choice models with an application to a store level reward program". mimeo. University of Toronto.

[28] Christensen, B.J. and N.M. Kiefer (1991). "The Exact likelihood function for an empirical job search model". *Econometric Theory* 7: 464-486.

[29] Christensen, B.J. and N.M. Kiefer (2009). *Economic Modeling and Inference*. Princeton: Princeton University Press.

[30] Classen, K.P (1977). "The effect of unemployment insurance on the duration of unemployment and subsequent earnings". *Industrial and Labor Relations Review* 30: 438-444.

[31] Cogan, J. (1981). "Fixed costs and Labor Supply". *Econometrica* 49: 945-964.

[32] Cohen-Goldner, S. and Z. Eckstein (2008): "Labor mobility of immigrants: Training, experience, language and opportunities". *International Economic Review* 49: 837-874.

[33] Cunha, F., Heckman, J. and S. Schennach (2006): "Estimating the technology of cognitive and noncognitive skill formation," mimeo.

[34] Daula, T. and R. Moffitt (1995). "Estimating dynamic models of quit behavior: The case of military reenlistment". *Journal of Labor Economics* 13, 499-523.

[35] Danforth, J.P. (1979). "On the role of consumption and decreasing absolute risk aversion in the theory of job search". in: S.A. Lippmand and J.J. McCall ed., *Studies in the economics of search*. New York: North Holland: 109-131.

[36] A.P. Dempster, M.Laird and D.B. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society* B 39: 1-38.

[37] Diamond, P.A. (1970). "A model of price adjustment". *Journal of Economic Theory* 3: 156-168.

[38] Diamond, P.A.and E. Maskin (1979). "An equilibrium analysis of search and breach of contracts I: Steady states". *Bell Journal of Economics* 10: 282-316.

[39] Dominitz J. and C.F. Manski. (1996) "Eliciting student expectations of the returns to schooling". *Journal of Human Resources* 31: 1–26.

[40] Dominitz J. and C.F. Manski (1997). "Using expectations data to study subjective income expectations". *Journal of the American Statistical Association* 92: 855–67.

[41] Eckstein, Z. and E. Nagypal (2004). "The evolution of US earnings inequality: 1961-2002". *Federal Reserve Bank of Minneapolis Quarterly Review* 28: 10-29.

[42] Eckstein, Z. and G. Vandenberg (2007). "Empirical labor search: A survey." *Journal of Econometrics* 136: 531-564.

[43] Eckstein, Z. and K. I. Wolpin (1989). "The specification and estimation of dynamic stochastic discrete choice models". *Journal of Human Resources* 24: 562-98.

[44] Eckstein, Z. and K. I. Wolpin (1989). "Dynamic labor force participation of married women and endogenous wage growth". *Review of Economic Studies* 56: 375-90.

[45] Eckstein, Z. and K. I. Wolpin (1990). "Estimating a market equilibrium search model from panel data on individuals". *Econometrica 58*: 783-808.

[46] Eckstein, Z. and K. I. Wolpin (1995). "Duration to first job and the return to schooling: estimates from a search-matching model". *Review of Economic Studies* 62: 263-286.

[47] Eckstein, Z. and K. I. Wolpin (1999). "Why youths drop out of high school: The impact of preferences, opportunities, and abilities". *Econometrica* 67: 1295-1339

[48] Eppel, D., R. Romano and H. Sieg (2006). "Admission, Tuition, and Financial Aid Policies in the Market for Higher Education." *Econometrica* 74: 885-928

[49] Ferrall, C. (1997). "Unemployment insurance eligibility and the school to work transition in Canada and the United States". *Journal of Business Economics and Statistics* 15: 115-129.

[50] Flinn, C. J. (2006) "Minimum wage effects on labor market outcomes under search, matching and endogenous contact rates". *Econometrica* 74: 1013-1062.

[51] Flinn, C. J. and J. J. Heckman (1982). "New methods for analyzing structural models of labor force dynamics". *Journal of Econometrics* 18: 114-142.

[52] Francesconi, M. (1998). "A joint dynamic model of fertility and work of married women". *Journal of labor Economics* 20: 336-380.

[53] Friedman, M. and S. Kuznets (1945). *Income from Independent Professional Practice.* New York: NBER.

[54] Fu, Chao (2009). "An Equilibrium Model of the College Market: An Empirical Investigation of Tuitopn, Applications, Admissions and Enrollment" mimeo, University of Pennsylvania.

[55] Gottschalf, P. and R. Moffitt (1994): "The growth of earnings instability in the US labor market". *Brookings Papers on Economic Activity* 2: 217-72.

[56] Gotz, G. A. and J. J. McCall (1984). "A dynamic retention model for air force officers: Theory and estimates," RAND, R-3028-AF.

[57] Gould, Eric (2008). "Marriage and career: The dynamic decisions of young men". *Journal of Human Capital* 2: 337-378.

[58] Greenberger E. and L. Steinberg ((1986). *When Teenagers Work: The Psychological and Social Costs of Adolescent Employment.* New York: Basic Books.

[59] Heckman, J. J. (1974). "Shadow prices, market wages and labor supply". *Econometrica* 42: 679-694.

[60] Heckman, J. J. (1979). "Sample selection bias as a specification error". *Econometrica* 47: 153-162.

[61] Heckman, J.J, L. Lochner and C.Taber (1998). "Explaining rising wage inequality: explorations with a dynamic general equilibrium model of earnings with heterogeneous agents." *Review of Economic Dynamics* 1: 1-58.

[62] Heckman, J.J and T. E. MaCurdy (1980). "A life cycle model of female labor supply". *Review of Economic Studies* 47: 47-74.

[63] Heckman, J.J and T. E. MaCurdy (1982). "Corrigendum on a life cycle model of female labor supply". *Review of Economic Studies* 49: 659-660.

[64] Heckman, J.J. and Sedlacek, G. (1985): "Heterogeneity, aggregation and market wage functions: An empirical model of self-selection in the labor market". *Journal of Political Economy*, 93: 1077-1125.

[65] Heckman, J.J. and B. Singer (1984). "A method for minimizing the impact of distributional assumptions in econometric models for duration data". *Econometrica* 52: 271-320.

[66] Heckman, J.J. and R.J. Willis. (1977) "A Beta-logistic model for the analysis of sequential labor force participation by married women". *Journal of Political Economy* 85:27-58.

[67] Hornstein, A, P. Krusell, and G.L. Violante (2005): "The effects of technical change on labor market inequalities," in: P. Aghion and S. Durlauf, eds., *Handbook of Economic Growth*, Amsterdam: Elsevier.

[68] Hotz, J. and R.A. Miller (1993). "Conditional choice probabilities and the estimation of dynamic models". *Review of Economic Studies* 60: 497-530.

[69] Hotz, J., R.A. Miller, S. Sanders and J. Smith (1994). "A simulation estimator for dynamic models of discrete choice". *Review of Economic Studies* 61: 265-289.

[70] Ichimura, H. and C. Taber (2002). "Semi-parametric reduced form estimation of tuition subsidies". *American Economic Association, Papers and Proceedings*: 286-292.

[71] Imai, S., N. Jain and A. Ching (2009). "Bayesian estimation of discrete choice models". *Econometrica* 77: 1865-1900.

[72] Imai, S. and M. P. Keane (2004). "Intertemporal labor supply and human capital accumulation". *International Economic Review* 45: 601-641.

[73] Imai, S. and K. Krishna (2004). "Employment, deterrence, and crime in a dynamic model". *International Economic Review* 45: 845-872.

[74] Jovanovic, B. (1979). "Job matching and the theory of turnover". *Journal of Political Economy* 87: 972-90.

[75] Juhn, C., K.M. Murphy and B. Pierce (1993). "Wage inequality and the rise in returns to skill". *Journal of Political Economy* 101: 410-42.

[76] Katz, L.F and D. H Autor (1999). "Changes in the age structure and earnings inequality". in: O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics* 3. Amsterdam: North Holland: 1463-1555.

[77] Katz, L. F and K. M Murphy (1992). "Changes in relative wages, 1963-1987: Supply and demand factors". *Quarterly Journal of Economics* 107: 35-78.

[78] Keane, M.P. (2009a). ""Labor supply and taxes: A survey". mimeo. University of Technology Sidney.

[79] Keane, M.P. (2009b). ""Income taxation in a life cycle model with human capital". mimeo. University of Technology Sidney.

[80] Keane, M. P. and R. Moffitt (1998). "A structural model of multiple welfare program participation and labor supply". *International Economic Review* 39: 553-590.

[81] Keane, M. P., R. Moffitt and D. Runkle (1988). "Real wages over the business cycle: estimating the impact of heterogeneity with micro data". *Journal of Political Economy* 96: 1232-1266.

[82] Keane, M.P. and R. M. Sauer (2009). "A computationally practical simulation estimator for panel data models with unobserved endogenous state variables". mimeo, University of Technology Sidney.

[83] Keane, M. P. and K. I. Wolpin (1994). "The solution and estimation of discrete choice dynamic programming models by simulation: Monte Carlo evidence". *Review of Economics and Statistics*: 648-672.

[84] Keane, M. P. and K. I. Wolpin (1997). "The career decisions of young men". *Journal of Political Economy* 105: 473-522.

[85] Keane, M. P. and K. I. Wolpin (2000). "Eliminating race differences in school attainment and labor market success". *Journal of Labor Economics* 18: 614-652.

[86] Keane, M. P. and K. I. Wolpin (2001). "The effect of parental transfers and borrowing constraints on educational attainment." *International Economic Review* 42: 1051-1103.

[87] Keane, M. P. and K. I. Wolpin (2002). "Estimating welfare effects consistent with forward looking behavior: An article in two Parts," *Journal of Human Resources* 37: 570-622.

[88] Keane, M. P. and K. I. Wolpin (2007). "Exploring the usefulness of a non-random holdout sample for model validation: Welfare effects on female behavior". *International Economic Review* 48: 1351-78.

[89] Keane, M. P. and K. I. Wolpin (forthcoming). "The role of labor and marriage markets, preference heterogeneity, and the welfare system in the life cycle decisions of black, Hispanic and white women". *International Economic Review.*

[90] Kiefer, N. M. and G.R. Neumann (1993). "Wage dispersion and homogeneity: The empirical equilibrium search model." in: H. Bunzel et. al. ed., *Panel Data and Labor Market Dynamics*. Amsterdam: North Holland.

[91] Kimmel J. and Knieser T.J. (1998). "New evidence on labor supply: employment vs. hours elasticities by sex and marital status". *Journal of Monetary Economics* 42: 289-301.

[92] Lee, D. (2005). "An estimable dynamic general equilibrium model of school, work and occupational choice". *International Economic Review* 46: 1-34.

[93] Lee, D. and K. I. Wolpin (2006). "Intersectoral labor mobility and the growth of the service sector". *Econometrica* 74: 1-46.

[94] Lee, D. and K. I. Wolpin (forthcoming). "Accounting for wage and employment changes in the U.S. from 1968-2000: A dynamic model of labor market equilibrium". *Journal of Econometrics.*

[95] Lentz, R. (2009). "Optimal unemployment insurance in an estimated job search model with savings", *Review of Economic Studies* 12: 37-57.

[96] McCall, J.J. (1970). "Economics of information and job search". *Quarterly Journal of Economics* 84: 113-126.

[97] MaCurdy, T.E. (1981). "An empirical model of labor supply in a life cycle setting". *Journal of Political Economy* 89: 1059-1085.

[98] MaCurdy, T.E. (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.

[99] MaCurdy, T.E. (1985). 'Interpreting empirical models of labor supply in an intertemporal framework with uncertainty". in J.J. Heckman and B. Singer eds. *Longitudinal analysis of labor market data.* Cambridge University Press: Cambridge U.K.

[100] Manski, C.F. (2004). "Measuring expectations". *Econometrica* 72:1329–76.

[101] Matzkin , R. L. (1993). "Nonparametric and Distribution-Free Estimation of the Binary Threshhold Crossing and the Binary Choice Models. Econometrica 60: 239-270.

[102] Mazzocco, M. and S. Yamaguchi (2006). "Labor supply, wealth dynamics and marriage decisions". Mimeo. UCLA.

[103] Miller, R. A. (1984). "Job matching and occupational choice". *Journal of Political Economy* 92: 1086-1120.

[104] Miller, R. A. (1997). "Estimating models of dynamic optimization with microeconomic data." in: M. H. Pesaran and P. Schmidt ed., *Handbook of Applied Econometrics: Microeconomics.* Basil Blackwell.

[105] Mincer, J. (1958). "Investment in human capital and personal income distribution". *Journal of Political Economy* 66: 281-302.

[106] Mincer, J. (1962). "Labor force participation of married women: A study of labor supply." in: *Aspects of Labor Economics.* Princeton, N.J.: National Bureau of Economic Research: 63-97.

[107] Moffitt, R. (1983). "An economic model of welfare stigma". *American Economic Review* 73: 1023-1035.

[108] Mortensen D. T. (1970) "A theory of wage and employment dynamics." in: E.S. Phelps et. al. ed., *Microeconomic Foundations of Employment and Inflations Theory*". New York: W.W. Horton.

[109] Mortensen D. T. (1982) "The matching process as a noncooperative game." in: J. J. McCall ed., *The Economics of Information and Uncertainty.* Chicago: NBER, University of Chicago Press.

[110] Mortensen D. T (1986). "Job search and labor market analysis" in O. Ashenfelter and R. Layard ed., *Handbook of Labor Economics.* Amsterdam: North-Holland.

[111] Murphy, K. M and F. Welch (1992). "The structure of wages". *Quarterly Journal of Economics* 107: 285-326.

[112] Murphy, K. M and F. Welch (1993). "Occupational change and the demand for skill: 1940-1990". *American Economic Review* 83: 122-126.

[113] Norets, A. (2009) "Inference in dynamic discrete choice models with serially correlated unobserved state variables". *Econometrica* 77: 1665-1682.

[114] Pagan, A. and A. Ullah (1999). *Nonparametric Econometrics.* Cambridge: Cambridge University Press.

[115] Pakes, A. (1986) "Patents as options: Some estimates of the value of holding European patent stocks". *Econometrica* 54: 755-84.

[116] Paserman, M. D. (2008) "Job search and hyperbolic discounting: Structural estimation and policy evaluation". *The Economic Journal* 118: 1418-1452.

[117] Postel-Vinay F. and J-M Robin (2002). "Equilibrium wage dispersion and with heterogeneous workers and firms". *Econometrica* 70: 2295-2350.

[118] Rendon, S.(2006). "Job search and asset accumulation under borrowing constraints". *International Economic Review* 47: 233-264.

[119] Rosen, S. (1977). "Human capital: A survey of empirical research," in: R. Ehrenberg ed., *Research in Labor Economics* 1. Greenwhich, CT: JAI Press.

[120] Roy, A. (1951). "Some thoughts on the distribution of earnings," Oxford Economic Papers 3: 135-146.

[121] Rust, J.)(1987) "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher". *Econometrica* 55: 999-1034.

[122] Rust, J. (1993). "Structural estimation of markov decision models." in: R. Engle and D. McFadden eds., *Handbook of Econometrics* Vol. 4., North Holland: Amsterdam.

[123] Rust, J. (1994). "Numerical dynamic programming in economics". in: H.M. Amman et. al. eds. *Handbook of Computational Economics.* North Holland: Amsterdam.

[124] Rust, J. (1997). "Using randomization to break the curse of dimensionality". *Econometrica* 65: 487-516.

[125] Sauer, R. (1998). "Job mobility and the market for lawyers". *Journal of Political Economy* 106: 147-171.

[126] Sauer, R. (2004). "Education financing and lifetime earnings". *Review of Economic Studies*71: 1189-1216.

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

[128] Stinebrickner, T. R. (2001). "A Dynamic model of teacher labor supply". *Journal of Labor Economics* 19: 196-230.

[129] Stern, S. (1989) "Estimating a simultaneous search model". *Journal of Labor Economics* 7: 348-369.

[130] Sullivan, Paul (forthcoming). "A dynamic analysis of educational attainment, occupational choices and job search". *International Economic Review.*

[131] Tartari, M. (2007). "Divorce and Cognitive Achievement of Children". Mimeo. Yale University.

[132] Todd, P. and K. I. Wolpin (2006). "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility". *American Economic Review* 96: 1384-1417.

[133] Todd, P. and K. I. Wolpin (forthcoming). "Ex-Ante Evaluation of Social Programs". *Annals of Economics and Statistics.*

[134] Van den Berg G.J. (1990). "Nonstationarity in job search theory". *Review of Economic Studies* 57: 255-277.

[135] Van den Berg G.J. and G. Ridder (1998). "An empirical equilibrium search model of the labor market". *Econometrica* 66: 1183-1221.

[136] Van der Klaauw, W. (1996). "Female labour supply and marital status decisions: a life cycle model". *Review of Economic Studies* 63: 199-235.

[137] Van der Klaauw, W. (2000). "On the use of expectations data in estimating structural dynamic models: an analysis of career choices". mimeo. University of North Carolina.

[138] Van der Klaauw, W. and K.I. Wolpin. (2008) "Social security and the retirement and savings behavior of low income households." *Journal of Econometrics* 145: 21-42.

[139] Van Soest, A. I. Woittiez and A. Kapteyn (1990). "Labor supply, income taxes and hours restrictions in the Netherlands". *Journal of Human Resources* 25: 517–558.

[140] Weiss, Y. and R. Gronau (1981). "Expected interruptions in labour force participation and sex-related differences in earnings". *Review of Economic Studies* 48: 607-619.

[141] Walsh, J.R. (1935): "Capital Concept Applied to Man." *Quarterly Journal of Economics* 49: 255-285.

[142] Welch, F. (1977). "What have we learned from empirical studies of unemployment insurance?" *Industrial and Labor Relations Review* 30: 450-461.

[143] Willis, R. and S. Rosen (1979): "Education and self-selection". *Journal of Political Economy* 87: S7-S36.

[144] Wise, D. (1985). "A behavioral model versus experimentation: The effects of housing subsidies on rent". in: P. Brucker and R. Pauly, eds., *Methods of Operations Research* 50: 441-489.

[145] Wolinsky, A. (1987). "Matching, search and bargaining". *Journal of Economic Theory* 42: 311-333.

[146] Wolpin, K. I. (1984). "A dynamic stochastic model of fertility and child mortality". *Journal of Political Economy* 92: 852-74.

[147] Wolpin, K. I. (1987). "Estimating a structural search model: The transition from school to work". *Econometrica* 55: 801-17.

[148] Wolpin, K. I. (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.

[149] Wolpin, K. I. (1995). *Empirical Methods for the Study of Labor Force Dynamics.* Harwood Academic Publishers: Luxembourg.

[150] Wolpin, K. I. (2003). "Wage Equations and Education Policy". in: M. Drewatripont et.al., Advances in Econometrics and Economics Vol. II. Cambridge University Press: Cambridge.

Figure 1: Hours, Wages and Price of Time over the Life-Cycle

Hours, Wage

Wage + HC

Wage

Hours

HC

Age
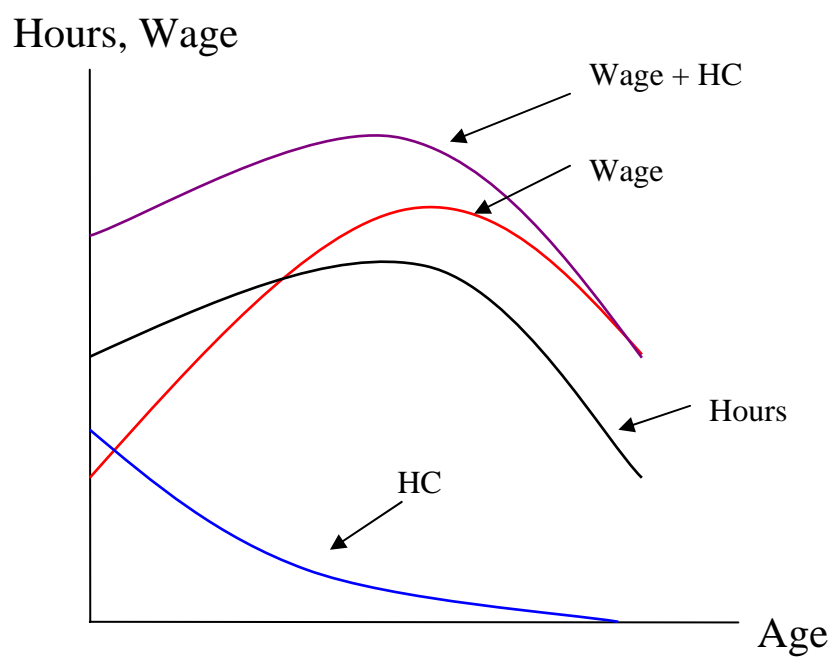
Note: HC denotes the return to an hour of work experience, in terms of increased present value of future wages. The opportunity cost of time is Wage + HC.

Table 1
Employment and Fertility of Married (White) Women: NLSY79

	Employment Hours (Ordered Probit)	Fetility (Probit)
Work Experience (Hours)	4.09 E-05	8.32E-06
	$(3.22\text{E-06})^3$	(4.33E-06)
Hours(t-1) = 1	1.04	-0.047
	(0.042)	(0.051)
Hours(t-1) = 2	1.90	-0.126
	(0.049)	(0.051)
Hours(t-1)=3	3.16	-0.222
	(0.110)	(0.089)
Age	-0.075	0.211
	(0.008)	(0.035)
Age Squared	-	-0.004)
		(0.0005)
Birth (t-1)	-0.497	-0.320
	(0.047)	(0.778)
Births(t-1 to t-5)	-0.349	0.448
	(0.031)	(0.054)
Total Births	0.099	-0.337
	(0.028)	(0.061)
Schooling	0.077	0.004
	(0.009)	(0.011)
Age of Spouse	0.007	-0.016
	(0.004)	(0.004
Schooling of Spouse	-0.036	0.021
	(0.007)	(0.010)
Marital Duration	-0.025	-0.015
	(0.006)	(0.008)
Constant	-	-3.41
		(0.497)
Cut point 1	-0.888	-
	(0.171)	
Cut point 2	0.076	-
	(0.172)	
Cut point 3	2.48	-
	(0.175)	
Pseudo R2	.295	.094

1. 8,183 person-period observations.
2. 8,786 person-period observations.
3. Robust standard errors in parenthesis.

Table 2
Effect of a 10% Tax on Earnings on Labour Supply by Age

Age	Pure Tax	Tax plus Lump Sum Redistribution
20	-0.7%	-3.2%
30	-0.7%	-3.3%
40	-0.9%	-4.2%
45	-1.2%	-5.7%
50	-2.1%	-8.7%
60	-9.1%	-20.0%
20-65 (Total Hours)	-2.0%	-6.6%