Review of Economic Dynamics 12 (2009) 1–22

Contents lists available at ScienceDirect

Review of Economic Dynamics

www.elsevier.com/locate/red

Empirical applications of discrete choice dynamic programming models

Michael P. Keane [a], Kenneth I. Wolpin [b,*,1]

[a] ARC Federation Fellow, University of Technology Sydney, Sydney, Australia
[b] University of Pennsylvania, PA, USA

ARTICLE INFO

Article history:
Received 14 January 2008
Revised 7 July 2008
Available online 12 July 2008

JEL classification:
C5
B41

Keywords:
Discrete choice dynamic programming models
Policy evaluation
Structural estimation

ABSTRACT

The development over the past 25 years of methods for the estimation of discrete choice dynamic programming (DCDP) models opened up new frontiers for empirical research in a host of areas, including labor economics, industrial organization, economic demography, health economics, development economics, political economy and marketing. In this paper, we first describe the development of the DCDP framework, showing how it was a natural extension of static discrete choice modeling. We then summarize six papers that adopt the DCDP paradigm that address substantively important social and economic questions. Finally, we consider the issue of the credibility of empirical findings based on the structural estimation of DCDP models.

1. Introduction[2]

The development of methods for the estimation of discrete choice dynamic programming (DCDP) models, that began over 20 years ago, opened up new frontiers for empirical research in a host of areas, including labor economics, industrial organization, economic demography, health economics, development economics and political economy, and has spread to areas outside of traditional economics, such as marketing.[3] There are a number of survey papers that describe the methodology and provide examples of applications found in the literature up to the mid-1990s (Eckstein and Wolpin, 1989a; Miller, 1997; Rust, 1993, 1994), and a recent survey that describes the methodological work that has been ongoing since that time (Aguirregabiria and Mira, in press). Another conventional survey would thus have limited value. The purpose of this essay is, instead, to provide evidence about whether the methodology's promise has borne fruit as a tool for empirical research. The success of what has come to be called, for better or worse, the "structural" approach to microeconomic empirical research should be judged according to the value of the empirical work that the approach has made possible.[4]

To that end, we describe how the DCDP paradigm has been used to address quantitatively six diverse, substantively important and challenging questions. The six papers that we summarize span the fields of economic demography, economic development, labor economics, health economics and political economy. The papers fall into two categories in terms of their goals. One of the papers attempts to explain a set of historical phenomena. Papers of this kind, to the extent that

* Corresponding author at: University of Pennsylvania, Department of Economics, 3718 Locust Walk, Philadelphia, PA, USA.
E-mail address: wolpink@ssc.upenn.edu (K.I. Wolpin).

1 Support from NSF grant SES-0450418 is gratefully acknowledged.

2 This essay derives from the plenary talk by Wolpin at the June 2007 SED conference in Prague.

3 Indeed, the DCDP approach is now arguably the predominant one in marketing.

4 The structural approach encompasses continuous as well as discrete choice models and static as well as dynamic models. Explicit modeling of dynamical considerations does not alter the basic rationale for the imposition of structure as contained in Marschak (1953).

1094-2025/$ – see front matter
doi:10.1016/j.red.2008.07.001

they are successful, provide models that improve our ability to forecast the future course of related events. The five other papers perform *ex ante* policy evaluations that attempt to assess the impact of policies outside of the historical experience. Papers of this kind, to the extent that they are successful, provide policy makers with the ability to compare the efficacy of alternative prospective policies.

The questions these six papers address are:

1. What are the economy-wide changes in the US since the mid 1960s that have been responsible for the increase in wage inequality, the increase in the college premium, the closing of the male–female wage gap, the increase in the labor force participation of women and the growth of the service sector?

2. To what extent can monetary incentives reduce racial differences in school attainment and wages?

3. What government policies would be most effective in increasing educational attainment in developing countries?

4. How will the new expansion of Medicare to incorporate prescription drug coverage affect the health and life expectancy of the elderly?

5. How would the career decisions of politicians be influenced by such policies as term limits, changes in Congressional salaries, and changes in post-Congressional career options?

6. What would be the impact of employer sanctions and increased border protection on the flow of illegal Mexican immigrants to the US?

We do not provide a critical assessment of the individual papers, but rather offer an overall picture of the strengths and weakness of the DCDP approach. As even the casual reader will recognize, the modeling assumptions in the individual papers used to address these questions are many and there is little doubt that other researchers would have chosen differently. One should not then, as we do not, think of these studies as in any way definitive. There is no such thing in empirical work. What we hope is that the reader will be convinced that these exercises comprise serious attempts to answer exceedingly difficult questions, that they make progress in doing so, and that their results should be taken at least as seriously as those derived from other empirical approaches. Although it goes without saying, these are not the only papers that tackle important questions using the DCDP approach, and many other papers would have equally well served our purpose.

We proceed first by describing the development of the DCDP framework. We show that it was a natural extension of the latent variable formulation of static discrete choice modeling. We begin, in Section 2, with the binary choice setting in order to fix ideas, using a simple labor force participation model as the prime motivating example. The DCDP framework is shown to have the same empirical structure as the static framework and, in that sense, raises no new conceptual estimation problems. Similarly, the computational issues that arise in the estimation of static multinomial choice models carry over to the dynamic case, but with the added complexity of having to numerically solve a dynamic programming problem. Section 3 summarizes the model structure and empirical results of the six papers that tackle the questions raised above. The value of the empirical work depends on the credibility that is attached to the findings. The next section addresses that issue, examining alternative methods of model validation. We include this section because practitioners of the DCDP approach themselves have been notably sensitive to it and because it is probably the most fundamental and vexing problem faced by empirical researchers of all stripes. The last section attempts to draw general lessons from the papers we review.

2. The common empirical structure of static and dynamic discrete choice models

The development of the DCDP empirical framework was a straightforward and natural extension of the static discrete choice empirical 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, denoted by i, makes a choice at each discrete period t, from $t=1$ to T, between two alternatives $d_{it} \in \{0,1\}$. Examples might be the choice of whether to work or not, whether (to try) to have a child or not, or whether to attend college or not. The outcome is determined by whether a latent variable, v^*_{it}, reflecting the difference in the payoff between choosing $d_{it}=1$ and $d_{it}=0$, 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 $v^*_{it} \geqslant 0$ and zero otherwise.

In its most general form, the latent variable may be a function of three types of variables: $\widetilde{D}_{it}$ is a vector of the history of all past choices ($d_{i\tau}\colon\ \tau=1,\ldots,t-1$), $\widetilde{X}_{it}$ is a vector of contemporaneous and lagged values of J additional variables ($X_{ij\tau}\colon\ j=1,\ldots,J;\ \tau=1,\ldots,t$) that enter the decision problem, and $\tilde{\epsilon}_{it}$ ($\epsilon_{i\tau}\colon\ \tau=1,\ldots,t$) is a vector of contemporaneous and lagged unobservables that also enter the decision problem.[5] The agent's decision rule at each age is given by

$$d_{it}=\begin{cases}1 & \text{if } v^*_{it}(\widetilde{D}_{it},\widetilde{X}_{it},\tilde{\epsilon}_{it}) \geqslant 0,\\ 0 & \text{if } v^*_{it}(\widetilde{D}_{it},\widetilde{X}_{it},\tilde{\epsilon}_{it}) < 0.\end{cases} \tag{1}$$

All empirical binary choice models are special cases of this formulation. Static models are distinguished from dynamic models by whether past choices affect the latent variable and, thus, the current choice. The goal of researchers

[5] 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.

is to estimate the (parameters of the) v_{it}^* function, where v_{it}^* itself is not observed.[6] There have been two approaches to building an empirically tractable model. In one approach, the latent variable function is interpreted as an "approximation" to the decision rule of a general class of optimization problems (Heckman, 1981; Wolpin, 1984; Keane and Wolpin, 2002).[7] The parameters of the latent variable function can be viewed as composites of the parameters of any optimization model from within the general class, i.e., as "reduced form" parameters. The second "structural" approach specifies and solves an explicit constrained optimization problem. Parameters are identified with specific structural relationships that led to the latent variable function. DCDP models fall within this approach. Their introduction is associated with independent contributions by Gotz and McCall (1984), Miller (1984), Pakes (1986), Rust (1987) and Wolpin (1984, 1987).

The basic insight of the structural approach to estimation is that any DCDP model can be cast as a static estimation problem. To illustrate the nature of that insight, it is useful to consider an example. We start with the simple static model of the labor force participation decision of a married woman, which has a long history in labor economics.[8]

2.1. Static models

2.1.1. Married woman's labor force participation

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

$$U_{it} = C_{it} + \alpha_{it}(1 - d_{it}), \tag{2}$$

where C_{it} is household i's consumption at period t and $d_{it} = 1$ if the wife works and is equal to zero otherwise. The utility the couple attaches to the wife's leisure (home production) is time-varying. Specifically,

$$\alpha_{it} = x_{it}\beta + \epsilon_{it} \tag{3}$$

where the variables in x_{it} $(\subseteq X_{it})$ include, among other things, the number of young children in the household, n_{it}, and where ϵ_{it} is a serially uncorrelated random shock to the utility of the wife's leisure. 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 child-care cost, π. The household budget constraint is thus

$$C_{it} = y_{it} + w_{it}d_{it} - \pi n_{it}d_{it}. \tag{4}$$

By substituting (3) and (4) into (2), we can write alternative-specific utilities, U_{it}^1 if the wife works and U_{it}^0 if she does not, as

$$\begin{aligned} U_{it}^1 &= y_{it} + w_{it} - \pi n_{it}, \\ U_{it}^0 &= y_{it} + x_{it}\beta + \epsilon_{it}. \end{aligned} \tag{5}$$

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

$$v_{it}^*(x_{it}, w_{it}, n_{it}, \epsilon_{it}) = w_{it} - \pi n_{it} - x_{it}\beta - \epsilon_{it}. \tag{6}$$

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, w_{it}, x_{it}, n_{it} and ϵ_{it}, from the part of the state space observable to the researcher, Ω_{it}^-, that is, consisting only of w_{it}, x_{it} and n_{it}. Given (6), the value of ϵ_{it} at which the woman is indifferent between working and not working is $\epsilon_{it}^*(\Omega_{it}^-) = w_{it} - \pi n_{it} - x_{it}\beta$. Households with a value of ϵ_{it} below $\epsilon_{it}^*(\Omega_{it}^-)$ choose to have the woman work, while households with a value of ϵ_{it} above $\epsilon_{it}^*(\Omega_{it}^-)$ choose otherwise. Assuming that the elements of Ω_{it}^- are distributed independently of ϵ_{it}, that is, they are taken to be exogenous, the probability of choosing $d_{it} = 1$, conditional on Ω_{it}^-, is given by

$$\Pr(d_{it} = 1 \mid \Omega_{it}^-) = \int_{-\infty}^{\epsilon_{it}^*} \mathrm{d}F_\epsilon(\epsilon_{it} \mid \Omega_{it}^-) = \int_{-\infty}^{\epsilon_{it}^*} \mathrm{d}F_\epsilon(\epsilon_{it}), \tag{7}$$

where $F_\epsilon(\epsilon_{it})$ is the distribution function for ϵ_{it}, and $\Pr(d_{it} = 0 \mid \Omega_{it}^-) = 1 - \Pr(d_{it} = 1 \mid \Omega_{it}^-)$. Thus, the likelihood function for a cross section of $i = 1, \ldots, I$, couples observed at different periods, the product of the sample choice probabilities, is

$$L(\theta; x_{it}, w_{it}) = \Pi_{i=1}^{i=I} \Pr(d_{it} = 1 \mid \Omega_{it}^-)^{d_{it}} \Pr(d_{it} = 0 \mid \Omega_{it}^-)^{1-d_{it}}, \tag{8}$$

where θ is the vector of parameters (β, π, F_ϵ) to be estimated.

---

[6] In some cases, the payoff to one alternative is observed, or partially observed (see below).

[7] The class of models is often, if not usually, left unspecified.

[8] Modern empirical research on this issue began with Mincer (1962).

It would be unusual to have data on wage offers, that is, both wages that are accepted and those that are declined. Without such data, it is necessary to specify a wage offer function in order to proceed with estimation. Let wage offers be generated by

$$w_{it} = z_{it}\gamma + \eta_{it}, \tag{9}$$

where z_{it} ($\subseteq X_{it}$) would conventionally contain educational attainment and, to keep the model static at this point, "potential" work experience (age − education − 6), and where the i.i.d. wage shock has distribution function F_η. Substituting (9) into (6) yields the latent variable function

$$\begin{aligned} v^*_{it}(x_{it}, z_{it}, \epsilon_{it}, \eta_{it}) &= z_{it}\gamma - \pi n_{it} - x_{it}\beta + \eta_{it} - \epsilon_{it} \\ &= \xi^*_{it}(\Omega^-_{it}) + \xi_{it}, \end{aligned} \tag{10}$$

where $\xi_{it} = \eta_{it} - \epsilon_{it}$, $\xi^*_{it}(\Omega^-_{it}) = z_{it}\gamma - \pi n_{it} - x_{it}\beta$ and Ω^-_{it} now consists of z_{it}, n_{it} and x_{it}.[9] The likelihood function incorporates the wage information, namely

$$L(\theta; x_{it}, z_{it}) = \Pi_{i=1}^{i=I} \Pr(d_{it} = 1, w_{it} \mid \Omega^-_{it})^{d_{it}} \Pr(d_{it} = 0 \mid \Omega^-_{it})^{1-d_{it}} \tag{11}$$

where $\Pr(d_{it} = 1, w_{it} \mid \Omega^-_{it}) = \Pr(\xi_{it} \geqslant -\xi^*_{it}(\Omega^-_{it}),\ \eta_{it} = w_{it} - z_{it}\gamma)$ and $\Pr(d_{it} = 0 \mid \Omega^-_{it}) = \Pr(\xi_{it} < -\xi^*_{it}(\Omega^-_{it}))$.[10]

It is unnecessary, for our purpose, to consider identification of this model in a semi-parametric setting, that is, without making distributional assumptions about the underlying unobservables (ϵ_{it} and η_{it}). To proceed, then, assume that $f(\epsilon, \eta)$ is joint normal with variance–covariance matrix, $\Lambda = \begin{pmatrix} \sigma^2_\epsilon & \cdot \\ \sigma_{\epsilon\eta} & \sigma^2_\eta \end{pmatrix}$. The parameters to be estimated include β, γ, π, σ^2_ϵ, σ^2_η, and $\sigma_{\epsilon\eta}$. Note first that it is not possible to separately identify the β associated with n_{it} in the utility function from the child care cost π. In the rest of the presentation, we normalize π to zero. Joint normality is sufficient to identify the wage parameters (γ and σ^2_η) as well as $(\sigma^2_\eta - \sigma_{\epsilon\eta}))/\sigma_\xi$ (Heckman, 1979). 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 x, that is, only in the value of leisure function, and variables that appear in both x and z. Having identified the parameters of the wage function (the γ'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. In the example, either education or potential experience must not affect the value of leisure.

Why should we impose this additional restriction? We can, after all, estimate how the participation probability varies with the x's and z's without parametric assumptions. There are two reasons why identifying structure is useful. First, 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 opportunities. Such an assessment could be useful in the design of public policies aimed at ameliorating those differences. Second, estimating the structure allows us to perform policy experiments that would otherwise not be possible. For example, suppose a policy maker was 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 (recalling that is π is normalized to zero) is

$$C_{it} = w_{it}d_{it} + y_{it} + \tau d_{it}n_{it}. \tag{12}$$

With the subsidy, the probability that the woman works is

$$\mathrm{pr}(d_{it} = 1 \mid z_{it}, x_{it}, n_{it}) = \Phi\big(z_{it}(\gamma/\sigma_\xi) - x_{it}(\beta/\sigma_\xi) + (\tau/\sigma_\xi)n_{it}\big), \tag{13}$$

where Φ is the standard normal cumulative. As seen in (13), it is necessary to have identified σ_ξ in order 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. Government outlays on the program would be equal to the subsidy amount times the number of women with young children who work (the number of married woman in the population x the estimated proportion that work as given by (13)).

Suppose that a researcher were to perform the desired policy experiment using the above model exactly as specified. We would conjecture that such an exercise would be met with skepticism. While we would like to see how robust the results are to alternative functional forms and distributional assumptions, the main objection, we believe, would be that the model is just too simple to be taken seriously. By too simple, we mean that there are too many aspects of the labor force participation decision that are being ignored, aspects that it would be reasonable to suppose might affect the estimate of the policy's impact.

[9] The additive error (ξ_{it}) is convenient in calculating choice probabilities and is maintained for illustrative purposes. However, the additive structure is fragile. It is lost, for example, if the wage function, as is conventional, takes on the semi-log form or if the utility function is non-linear in consumption.

[10] The Jacobian of the transformation from the wage density to the wage error density is one.

2.2. Dynamic models

In the static model, there is no connection between the decision made in the current period and future utility. Thus, even if agents are forward looking, maximizing the expected present value of discounted lifetime utility would be equivalent to maximizing current utility in each period. There are many ways in which dynamics may arise in the two models. For example, suppose that for the labor force participation problem, the woman's wage increases with actual work experience, h. In that case, rewrite (9) as

$$w_{it} = z_{it}\gamma_1 + \gamma_2 h_{it} + \eta_{it}, \tag{14}$$

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 wages. Or, in the child labor case, suppose, for example, that parents' utility depends not only on current attendance, but also on completed schooling, the sum of years of school attendance, by some fixed terminal school leaving age.

Continuing with the labor force participation example, modified to account for the wage offer function in (14) but otherwise maintaining the same structure, assume that the couple maximizes the expected present discounted value of remaining lifetime utility at each period starting from an initial period, $t = 1$ and ending at period T, the assumed terminal decision period.[11] Letting $V_t(\Omega_{it})$ be the maximum expected present discounted value of remaining lifetime utility at t given the state space and discount factor δ,

$$V_t(\Omega_{it}) = \max_{d_{it}} E\left\{\sum_{\tau=t}^{\tau=T} \delta^{\tau-t}\left[U_{it}^1 d_{it} + U_{it}^0(1-d_{it})\right] \mid \Omega_{it}\right\}. \tag{15}$$

The state space at t consists of all factors, known to the individual at t, that affect current utility or the probability distribution of future utilities. With the wage equation given by (14), h_{it} becomes part of the state space and evolves according to

$$h_{it} = h_{i,t-1} + d_{i,t-1}. \tag{16}$$

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

$$V_t(\Omega_{it}) = \max\bigl(V_t^0(\Omega_{it}), V_t^1(\Omega_{it})\bigr), \tag{17}$$

each of which obeys the Bellman equation

$$V_t^k(\Omega_{it}) = \begin{cases} U_{it}^k + \delta E[V_{t+1}(\Omega_{i,t+1}) \mid \Omega_{it},\ d_{it} = k] & \text{for } t < T, \\ U_{iT}^k & \text{for } t = T. \end{cases} \tag{18}$$

The expectation in (18) is taken over the distribution of the random components of the state space at $t+1$ conditional on the state space elements at t, i.e., over the unconditional joint distribution of the random shocks, given that all shocks are mutually serially independent.

The latent variable in the dynamic case is the difference in alternative specific value functions, $V_t^1(\Omega_{it}) - V_t^0(\Omega_{it})$, namely

$$\begin{aligned} v_t^*(\Omega_{it}) &= z_{it}\gamma_1 + \gamma_2 h_{it} - x_{it}\beta - \epsilon_{it} + \eta_{it} \\ &\quad + \delta\bigl\{\bigl[E\bigl[V_{t+1}(\Omega_{i,t+1}) \mid \Omega_{it}, d_{it} = 1\bigr]\bigr] - \bigl[E\bigl[V_{t+1}(\Omega_{i,t+1}) \mid \Omega_{it}, d_{it} = 0\bigr]\bigr]\bigr\} \\ &= \xi_{it}^*(\Omega_{it}^-) + \xi_{it}. \end{aligned} \tag{19}$$

Comparing the latent variable function in the dynamic (19) and static (10) cases, the only difference is the appearance of the difference in the future component of the expected value functions under the two alternatives.[12] A full solution of the dynamic programming problem consists of finding $E[\max(V_t^0(\Omega_{it}), V_t^1(\Omega_{it}))]$ at all values of Ω_{it}^-, denoted by $E\max(\Omega_{it}^-)$, for all $t = 1,\ldots,T$. Given such a solution, as seen in the second equality in (19), the estimation of the dynamic model is in principle no different than the estimation of the static model. The only conceptual difference is that the dynamic problem introduces an additional parameter, the discount factor, δ. The practical difference in terms of implementation may be large.

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 and denote t_{1i} and t_{Li} are the first and last periods of data observed for individual i. In that case, the likelihood function is

$$L(\theta; x_{it}) = \Pi_{i=1}^{i=I}\Pi_{\tau=t_{1i}}^{\tau=t_{Li}} \Pr(d_{i\tau} = 1, w_{i\tau} \mid \Omega_{it}^-)^{d_{i\tau}} \Pr(d_{i\tau} = 0 \mid \Omega_{it}^-)^{1-d_{i\tau}}, \tag{20}$$

---

[11] Eckstein and Wolpin (1989b) empirically implemented this model allowing for additional sources of dynamic behavior. The finite horizon assumption is immaterial for the points we wish to make.

[12] We have substituted actual experience in z_{it} for potential experience, although including potential experience as well would be identical to including age.

where $\Pr(d_{i\tau}=1, w_{i\tau}) = \Pr(\xi_{i\tau} \geqslant -\xi^*_{i\tau}(\Omega^-_{i\tau}),\ \eta_{i\tau} = w_{i\tau} - z_{i\tau}\gamma_1 - \gamma_2 h_{it})$ and $\Pr(d_{i\tau}=0) = 1 - \Pr(\xi_{i\tau} \geqslant -\xi^*_{i\tau}(\Omega^-_{i\tau}))$.[13] Note that the difference in the future component of the expected value functions under the two alternatives in (19) is a non-linear function of the state variables, z_{it}, h_{it}, x_{it}, and depends on the same set of parameters as in the static case. Rewriting (19) as

$$v^*_t(\Omega_{it}) = z_{it}\gamma_1 + \gamma_2 h_{it} - x_{it}\beta + \delta G(z_{it}, h_{it}, x_{it}) - \epsilon_{it} + \eta_{it}, \tag{21}$$

where $G(\cdot)$ is the difference in the future component of the expected value functions, the non-linearities in G that arise from the distributional and functional form assumptions may be sufficient to identify the discount factor. Identification requires the same exclusion restriction as in the static case, that is, the appearance of a 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.[14]

2.2.1. The multinomial choice problem

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 it was for the binary choice problem. In fact, it does not do much injustice to simply allow the number of mutually exclusive alternatives, and thus the number of alternative-specific value functions in (18) 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 $E[\max(V^0_t(\Omega_{it}), V^1_t(\Omega_{it}), \ldots, V^K_t(\Omega_{it}))]$ function that enters the multinomial version of (18) at all values of Ω^-_{it} and at all t, adds significantly to that computational burden. It will be useful for the later exposition to define d^n_{it} as the discrete $\{0, 1\}$ choice variable corresponding to the nth choice $(n = 1, \ldots, N)$ and $\tilde{d}_{it}$ the N element vector of those choices. There would be, then, at most $K = 2^N$ mutually exclusive choices.

2.2.2. Unobserved heterogeneity

The stochastic components of the model, preference and wage shocks, were assumed to be mutually serially uncorrelated. There is nothing that rules out general serial correlation other than computational feasibility in solving the dynamic programming problem and in estimating the model. A standard specification that allows for serial dependence, conditional on observable state variables, assumes that agents can be distinguished, in terms of preferences and opportunities, by a fixed number of types. For example, in the labor force participation model with J types, if the woman was of type j, her preference for leisure might be specified as $\alpha_{ijt} = \alpha_{oj} + x_{it}\beta + \epsilon_{it}$ and her wage offer as $w_{ijt} = \gamma_{oj} + z_{it}\gamma_1 + \gamma_2 h_{it} + \eta_{it}$.[15] The type j woman would be distinguished by the $(\alpha_{oj}, \gamma_{oj})$ pair. Thus, potentially, women who value leisure more might also be more productive in the labor market. The dynamic program, in this case, must be solved for each of the J types and the likelihood function is a weighted average of the type-specific likelihoods. The weights are the proportions of each type in the sample and are estimated along with the other parameters.

2.2.3. More flexible specifications

Estimable DCDP models are, by necessity, fully parametric. However, any DCDP that can be numerically solved can, in principle, be estimated. The DCDP structure does not restrict functional forms for preferences, technologies and institutional constraints (e.g., tax rules), including the way in which unobservables enter (e.g., non-additive errors, serial correlation), nor does it restrict the distributional assumptions of unobservables. The restrictions arise from practical considerations, e.g., about the size of the state space and the number of parameters that must be estimated.

3. Selected applications of discrete choice dynamic programming models

In this section, we review six empirical applications of the DCDP approach. We have selected the papers to show the diversity of areas and questions to which the approach has been applied and with the view that these papers address substantively important and challenging social and economic issues. These papers also illustrate the variety of structures that the DCDP approach can accommodate.

---

[13] 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.

[14] It would be problematic, however, if the unobservables that determine either wages or the value of leisure are not i.i.d. over time.

[15] Although the example assumes that the heterogeneity is in the intercept terms, more general specifications in which slope coefficients are heterogeneous are also possible.

Table 1
Selected changes in hourly wages and employment: 1968–2000[a]

	(1) 1968–1974	(2) 1995–2000	(2) ÷ (1)
Hourly wage rate[b]			
90th percentile	15.88	21.10	1.33
10th percentile	4.10	4.24	1.03
Median			
College grad.	12.01	14.04	1.17
HS grad.	7.92	7.89	1.00
Males	10.02	10.93	1.09
Females	5.86	8.01	1.37
Employment rate[c]			
Males	0.721	0.728	1.01
Females	0.366	0.578	1.58
Service/Goods	1.58	2.81	1.78

[a] Ages 16 to 65.
[b] 1983 dollars.
[c] Annual hours 780 or more.

3.1. ***Question:*** *What are the economy-wide changes in the US since the mid 1960s that have been responsible for (i) increased wage inequality, (ii) the increase in the college premium, (iii) the closing of the male–female wage gap, (iv) the increased labor force participation of women, (v) the growth of the service sector?*
Donghoon Lee and Kenneth I. Wolpin (in press)

The US labor market looks very different today than it did 30 years ago. Wage inequality has increased, the wage return to a college degree has risen, the gap between the wages of working women and men has narrowed, employment rates of women have increased and a higher proportion of workers are found in the service sector of the economy. Table 1 summarizes these trends. The hourly wage of workers in the 90th percentile of the distribution grew by 33% between the periods 1968–1974 and 1995–2000, while that for workers in the 10th percentile grew only by 3%. Concomitantly, the wage of college graduates (16 or more years of schooling) grew by 17% over the same period while the wage of high school graduates (exactly 12 years of schooling) was unchanged. Over the period, the wages paid to women grew by 37% and to men by 9%, narrowing the female-to-male wage gap from 58 to 73%. At the same time, the fraction of women who worked increased by 58% while the fraction of men who worked remained roughly the same. Finally, in the 1968–1975 period, there were 1.58 persons employed in the service sector for every person employed in the goods-producing sector, but by the 1995–2000 period that number had jumped to 2.81.

Lee and Wolpin conjecture that these phenomena (and a number of others not considered here) are interrelated, resulting from the same set of changes in underlying fundamentals. Their paper builds on an extremely large literature that has sought to explain these trends, a literature that, however, has mostly considered them as separate phenomena. Katz and Autor (1999) provide a comprehensive review. As they point out, the primary framework for explaining these phenomena is based on standard demand and supply considerations. Among those considerations are:

Demand-side factors:

1. Factors that contribute to shifts towards increased employment in high skill industries—differential technological change across sectors.
2. Factors that contribute to skill upgrading within industries—skill biased technological change within sectors; capital–skill complementarity coupled with changes in the price of capital favoring capital-intensive sectors.
3. International trade changes—changes in relative product prices favoring skill-intensive sectors.

Supply-side factors:

1. Cohort size variation.
2. Factors that alter female labor supply given factor prices–fertility; changes in the value of home time or in social “norms” associated with female work.
3. Factors that alter educational investment behavior given factor prices.
4. Immigration.

The approach adopted by Lee and Wolpin differs from that found in the literature in at least two ways. First, they develop and structurally estimate an explicit labor market *equilibrium* model.[16] Second, they quantitatively assess the relative importance of a number of interrelated explanations for these phenomena, some that are potentially countervailing.[17] It is the first, the adoption of a structural approach, that permits the second. The general features of the model they develop and estimate are:

(i) There are two production sectors: the goods sector and the services sector. The aggregate production functions are CES with constant returns to scale in three skill types, white-, pink- and blue-collar occupations, and homogeneous capital. Capital and white-collar skill form a nested CES composite input (capital–skill complementarity). There are both time-varying neutral and non-neutral technological change and combined aggregate productivity shocks and real product price shocks. Changes in real goods and service product prices are exogenous as is the price of capital.[18]

(ii) Individuals age 16 to 65 choose among eight discrete and mutually exclusive alternatives at each age, working in any of the six sector-occupations, attending school or remaining in the home sector. An individual receives a stochastic wage offer from each sector-occupation in each period that is the product of a competitively determined sector-occupation-specific skill rental price and the individual's accumulated sector-occupation-specific skill. The latter depends on the individual's level of schooling in that period and the individual's accumulated work experience in each sector-occupation. There is an age-invariant non-pecuniary payoff to each sector-occupation and a stochastic consumption value of attending school and of remaining in the home sector. The latter depends on the number of pre-school children in the household and also is trended due to technological improvements in household production technology. Transiting between alternatives involves a cost. The model allows for gender differences in preferences, for example, in the value of the home alternative when there are young children in the household.

(iii) The population at any calendar time consists of overlapping generations of individuals of both sexes age 16 to 65 and is time-varying due to past fertility and immigration. The population consists of five types of individuals, where a type is distinguished by their endowments (at age 16) of each sector-occupation-specific skill and of the values they attach to attending school or remaining at home.

The model is solved for the six equilibrium skill rental prices, which are determined by equating skill prices to their respective marginal revenue products evaluated at aggregate skill amounts. Lee and Wolpin adopt a forecasting rule, that is, a joint stochastic process, for skill prices, and develop an iterative algorithm to determine the parameters of the process. They estimate the parameters of the model by matching aggregate statistics on employment, wages and school enrollment from the March supplements of the Current Population Surveys from 1968 to 2001, on sectoral output and capital from the Bureau of Economic Analysis and on employment transitions and wage growth from the NLSY79.

The model, given the parameter estimates, is used to simulate counterfactual experiments to quantitatively assess the relative importance of the supply and demand factors mentioned above. That assessment makes use of the following thought experiment. Suppose the world had stopped changing after 1960 in terms of demand and supply factors, for example, there was no further Hicks-neutral or skill-biased technological change, no further changes in fertility rates or in the value of leisure. Compared to that world (the 1960 base), how would the US labor market have evolved under alternative scenarios in which some of these factors changed as they did in actuality and others did not change, and how would those new worlds differ from what actually happened (as predicted by the model)? A selection of the results of this exercise are provided in Table 2 and can be summarized briefly as follows:

1. Wage inequality:

As seen in the baseline, between the 1968–1974 and 1994–2000 periods, the ratio of the 90th to the 10th wage percentile grew by 18% relative to the no-growth 1960 base. The first two counterfactual experiments, changes in supply factors, that is, in cohort size and in the number of young children in the household, and in the value of home time, indicate that those changes by themselves would have reduced wage inequality slightly. The third counterfactual that allows for Hicks-neutral technological change caused, by itself, a 5% increase in this measure of inequality and skill-biased technological change a 14% increase. In order for demand-side factors to induce increased wage inequality, the supply response in terms of sector-occupation skills must not be perfectly elastic. Although sector-occupation transition costs, the accumulation of sector-occupation-specific work experience and random preference and skill shocks contribute to the response inelasticity, the primary source of the skill supply inelasticity in the model is due to unobserved heterogeneity in skill endowments and in preferences. This "friction" shows up in changes in relative skill prices. For example, skill biased technological change leads to an increase of 35 (19) percent in the ratio of the white-collar to blue-collar skill rental price in the goods (service) sector.

2. The college wage premium

The growth in the college premium, 28% in the baseline, like the general increase in inequality, is estimated to have been overwhelmingly a technology driven outcome. Like wage inequality, skill-biased technological change was considerably more important, leading to a 41% increase in the premium, than was Hicks neutral change, which, by itself, would have led to a 21% increase. Heterogeneity, as above, accounts for the incomplete skill supply response.

[16] The model has recent antecedents in Keane and Wolpin (1997), Heckman et al. (1998), Lee (2005), and Lee and Wolpin (2006).

[17] An exception to the consideration of these as separate phenomena is Welch (2000).

[18] Hicks neutral technological change is not distinguished from changes in relative product prices.

Table 2
The relative impact of demand and supply factors on labor market changes in wages and employment from 1968 to 2000

	Baseline	Change in fertility and cohort size	Trend change in value of leisure	Hicks neutral technological change[a,b]	Skill-biased technological change[c]
Increase in wage inequality (Log 90/10)	18	−2	−5	5	14
Increase on college wage premium	28	−5	−4	21	41
Increase in female/male wage ratio	24	4	12	7	8
Increase in female/male employment ratio	84	42	24	21	23
Increase in service/goods sector employment ratio	77	8	16	73	10

[a] Includes real product price changes.
[b] 26% growth in goods sector and 102% growth in service sector real output from 1968–1974 to 1995–2000.
[a] 149% increase in MP of white-collar vs. blue-collar workers in goods sector, 72% change in service sector.

3. Female–male wages and employment
A number of factors played a role in the increased relative wages and employment of women. Changes in the value of home time, by itself, would have caused a 12% increase in the female-to-male wage ratio and a 24% increase in the female-to-male employment ratio. Changes in cohort size and fertility would have caused a 42% increase in relative employment. All of the supply side factors taken together (not in table) would have caused a 17% increase in relative wages and an 81% increase in relative employment. Dynamics are important. The fall in women's value of leisure that was estimated in the model to have occurred induces women to invest more heavily in market skills, given that they anticipate spending a larger proportion of their lifetime in the labor market. Thus, they obtain more schooling and more work experience, both of which raise their wage absolutely and relative to men. Thus, unlike in a static model, an increase in the supply of women to the workforce can lead to an increase in the relative wage of women. In fact, given that male and female skill within sector-occupations are perfect substitutes, Lee and Wolpin report that male wages actually would have fallen. Technological change would have also, by itself, caused a significant increase in the female-to-male relative wage, with Hicks neutral and skill-biased technological change having effects of similar magnitudes, of 7 and 8% respectively. They also would have caused similar increases in the relative employment of women, of 21 and 23% respectively.

4. Service–goods sector employment
The sectoral difference in Hicks-neutral technological change was the predominant factor in the relative growth in service sector employment, causing a 73% increase in the ratio of service-to-goods sector employment. Changes in the value of home time, which as seen would have induced women to enter the labor force, by itself, would have caused relative employment to change by 16%, larger than the effect of skill-biased technological change.

Was it important to allow for both supply and demand factors in explaining these labor market changes? The answer is mixed. Supply side factors contributed little to the change in overall wage inequality or to the college premium. The importance of technological change in accounting for these trends is consistent with the literature that has either ignored supply factors or have not integrated them in a single framework. On the other hand, while supply factors were quantitatively more important in explaining gender changes in wages and employment, demand factors by themselves would have led to non-negligible changes. Conversely, service sector growth, while predominantly a demand side phenomenon, was also importantly affected by the growth in female employment. Thus, in both of these latter cases, an important part of the picture would be missing in the absence of the kind of integrated empirical framework allowed by the structural DCDP approach.

3.2. ***Question:*** *To what extent can monetary incentives reduce racial differences in school attainment and wages?* *Michael P. Keane and Kenneth I. Wolpin (2000)*

The disparity in educational attainment between white and African American males is substantial. For example, Keane and Wolpin (2000) report that for a subset of males in the 1979 youth cohort of the NLSY who were 13–16 years old by October 1977 (people born in the early 1960s), mean highest grade completed at age 23 (at which point education is mostly completed) was 12.7 for whites compared to 11.9 for blacks. In addition, 21.3% of whites had completed college, compared to only 6.8% of blacks, and 35.9% of blacks were high school dropouts, compared to 24% of the whites.

Keane and Wolpin (2000) estimate a model of human capital investment decisions over the life-cycle to: (i) analyze the relative importance of different factors that drive these racial differences in educational attainment, (ii) determine the extent to which these differences can be ameliorated by financial incentives for black youth to continue in school, and (iii) analyze the influence of such incentives on subsequent labor market outcomes.

The model that Keane and Wolpin (2000) use is essentially identical to that in Keane and Wolpin (1997), "The Career Decisions of Young Men." In that paper, Keane and Wolpin develop a DCDP model where, in each annual period, a person has five mutually exclusive options: school attendance, white-collar work, blue collar work, military employment and "home." The payoff to the school option is the monetized value of the utility/disutility from attending school, net of tuition costs. The payoffs from the employment options are the wages earned as well as non-pecuniary rewards from working in particular occupations. The payoff to the home option is the monetized value of home production and/or leisure. The school and home payoffs are subject to additive shocks, while wages in the three employment options are subject to multiplicative shocks (additive shocks in the log earnings equations). Shocks to wages and non-pecuniary rewards are revealed in each period before agents decide among the five options, but future shocks are not yet known. Agents maximize the expected present value of lifetime earnings plus the monetized value of non-pecuniary rewards over a finite horizon, starting from age 16 and ending at age 65.

The most important source of dynamics in the model arises through the specification of the earnings functions. Earnings are determined by a (time invariant) rental rate on human capital times the stock of human capital. The human capital stock is augmented by education and by occupation-specific and overall work experience. Thus, current decisions affect future human capital and hence future wages. Human capital is, thus, multi-dimensional. Agents have distinct stocks of human capital that are relevant for each occupation, that is, separate stocks of white-collar human capital and of blue-collar human capital. According to Keane and Wolpin's estimates, education augments white-collar capital more than it augments blue-collar capital, while for work experience it is the reverse.

Dynamics also arise due to costs of moving between occupations. For instance, there is a fixed cost of moving into white- or blue-collar work if one was not working in that occupation in the preceding year. There is also a cost of returning to school if one had not been in school the previous year. Keane and Wolpin (1997) found that such mobility costs were important in order for the model to replicate the level of persistence in choices observed in the data, as well as the steepness of the decline of schooling with age.

A key element of the model is the existence of unobserved heterogeneity. Keane and Wolpin (1997) allow for four types of individuals. A type is distinguished by initial, age 16, white- and blue-collar "skill or human capital endowments," and also by tastes for school and home formed by age 16.[19] The type-specific parameters, as well as the population proportion of each type, are both estimated. Because each type has a vector of skill endowments and tastes, type distinctions are multidimensional. For instance, the "types 1" are estimated to have a high endowment of white-collar skill and enjoy school. Hence, they tend to follow a career trajectory where they complete college and then go into white-collar work. The "types 2" have high endowments of blue-collar skill, higher than those of type 1, and do not like school. Hence, they tend to follow a career trajectory where, after leaving high school, they quickly move into blue-collar jobs at relatively high wages. Types 3 and 4 have much lower skill endowments. But, they differ in that types 3 have a lower value of leisure. Thus, while both groups tend to have fairly low schooling levels, it is the types 4 who account for most of the high school dropouts and for those that choose the home sector. Types 3, in contrast, generally move into the military or blue-collar work shortly after leaving high school.

How does the model explain the very large differences in choices (i.e., school attendance, work, occupational choice) and in realized earnings for blacks and whites? To answer this question, Keane and Wolpin (2000) first estimated the model separately for blacks, that is allowing all of the parameters to differ. They then estimated a restricted version of the model in which all parameters for blacks and whites were constrained to be equal, except for (i) the type proportions and (ii) the intercepts in the earnings functions. The type proportions capture race differences in age 16 skill endowments and preferences through a shift in type-specific population weights. The intercepts capture both race-differences in skill rental prices due to discrimination and differences in the level of the skill endowments of each type.[20] Strikingly, this severe restriction on the model leads to a rather trivial deterioration in model fit. However, the structural differences between the races are large. Specifically, Keane and Wolpin (2000) estimate that the constant terms in the wage offer equations are 7.9% lower for blacks in the white-collar occupation and 4.7% lower in the blue-collar occupation. The differences in type proportions are even more pronounced: only 17.4% of blacks are types 1 or 2 (the high wage types), compared to 43.2% of whites. Conversely, 41.6% of blacks are types 4 (which accounts for most high school drop outs) compared to only 20.3% of whites.

The main point of the paper is to examine whether financial incentives for school attendance can reduce or even eliminate racial differences in school attainment and wages. As shown in Table 3, based on the model estimates, Keane and Wolpin calculate that a scheme that provided blacks with a bonus of $6250 for finishing high school and an additional $15,000 for finishing college (both in 1987 dollars) would essentially equalize the distribution of completed education for whites and blacks. They calculate that implementation of such a scheme would cost about $2 billion per year (in 1987 dollars).

One might naturally expect that the equalization of schooling would go a long way to equalizing lifetime earnings, both because schooling directly increases earnings in all occupations and because those with higher schooling work in high wage (white-collar) occupations. Surprisingly, however, Keane and Wolpin find that the large increase in education induced by this scheme has very little impact on the mean present value of lifetime earnings for blacks. Under the baseline, Keane and Wolpin simulate that the present value of lifetime earnings is $285,000 for whites and $210,000 for blacks, so lifetime

[19] "Tastes" for schooling and for home are isomorphic with productivities in those sectors.

[20] In a Ben-Porath (1967) human capital pricing equation, the interpretation of the log wage equation adopted by Keane and Wolpin, it is not possible to separately identify the skill rental price and the skill endowment from the constant term. Thus, Keane and Wolpin (2000) note that the race difference in the constant terms provides an upper bound estimate for the extent of discrimination.

Table 3
Black–white differences in education, earnings and utility

	NLSY79 data[a]		Baseline model[a]		School attendance bonus	Equalize type proportions	Equalize wage offer constants
	Whites	Blacks	Whites	Blacks	Blacks	Blacks	Blacks
Mean HGC	12.7	11.9	13.0	12.0	12.8	12.6	12.3
Completed school:							
<HS	24.0	35.9	26.2	37.9	26.0	34.1	32.7
High school	36.2	37.7	29.6	31.5	31.6	28.5	31.6
Some college	18.7	19.6	19.1	17.9	16.2	16.4	20.6
College+	21.3	6.8	25.1	12.8	26.1	21.1	15.3
Mean earnings:							
Age 30	na	na	22,795	15,818	16,526	19,987	18,327
PV of lifetime:							
Earnings	na	na	285,400	210,258	213,885	254,228	239,219
Utility	na	na	316,898	264,886	269,402	298,552	278,602

[a] The figures for completed schooling in the NLSY79 data are measured as of age 23. The figures for the baseline model are for the highest level of schooling ever attained in the simulation. The latter are slightly higher than the former.

earnings for blacks is only 73.7% that of whites. The school attendance bonus scheme increases lifetime earnings for blacks only up to $214,000, which is 75.1% of that for whites. Thus, the bonus scheme eliminates only about 5% of the gap.

Why are financial incentives for school attendance so effective at increasing attendance itself, but so ineffective at increasing earnings? The result that reasonably modest attendance bonuses would substantially increase college attendance by African Americans is consistent with a large literature that finds a rather elastic response of college attendance to college costs (for reviews, see, e.g., Leslie and Brinkman, 1987 or Keane and Wolpin, 2001). However, raising college attendance *per se* does nothing to alter the age 16 endowments or to reduce wage discrimination, the factors that account for the earnings differential between blacks and whites.

The importance of these factors can be seen through two hypothetical experiments that Keane and Wolpin (2000) conduct using their model. Specifically, as seen in Table 3, they find that equalizing type proportions (i.e., giving blacks the same type proportions as whites) would bring the present value of lifetime earnings for blacks up to $254,000, while eliminating racial discrimination and/or equalizing type-specific skill endowments would bring it up to $239,000. Thus, these hypothetical experiments would eliminate 52 and 39% of the earnings gap, respectively. These are large effects, particularly relative to the effects of equalizing school attainment.

The main conclusion that Keane and Wolpin (2000) draw from these results is that policies like college financial aid, school attendance bonuses, etc., can only be effective ways to reduce lifetime earnings inequality between blacks and whites if they also act to change age-16 endowments.[21] These endowments may be determined by many interrelated factors, such as parental education and income, parental investments in children, pre- and post-natal care, the quality of child care, pre-schools, primary schools and secondary schools, neighborhood effects, etc. The extent to which the kind of educational policies considered above induce behaviors that affect endowments is unknown. To develop effective policy, we must therefore get inside the black box of what determines age-16 skill endowments. There is now a growing research agenda that is focusing attention on the formation of both cognitive and non-cognitive skills during childhood.[22]

*3.3. **Question:** What government policies would be most effective in increasing educational attainment in developing countries? Petra E. Todd and Kenneth I. Wolpin (2006)*

It has been a central tenet of education policy in developing countries that human capital plays a critical role in economic growth and in income inequality. Given this view, developing countries have actively pursued innovative policies to increase educational attainment. In 1997, the Mexican government introduced a conditional cash transfer (CCT) program in rural areas that provided a subsidy to families for each child that met a school attendance goal. The initial program, called PROGRESSA, has since been extended to urban areas in Mexico (and renamed Opportunidades), and has been adopted in numerous other countries (for example, in Bangladesh, Brazil, Colombia, Guatemala, Pakistan).

To evaluate the program, the Mexican government conducted a randomized social experiment, in which 506 rural villages were randomly assigned to either participate in the program or serve as controls. Randomization, under ideal conditions, allows mean program impacts to be assessed through simple comparisons of outcomes for treatments and controls. The program was effective in increasing school attendance; treatment effects, measured as the difference in average attendance

[21] Keane and Wolpin (2000) also used their model to examine whether wage subsidies for low wage workers would be effective as a means to reduce earnings inequality. Unfortunately, such policies have the perverse effect of inducing low skill endowment workers to optimally choose even less schooling.

[22] Although the early human capital literature recognized the importance of understanding skill formation in young children (Leibowitz, 1974), the human capital literature became focused, overly so in our view, on estimating the impact of years of school attainment on wages (Wolpin, 2003). The recent revival of interest in early childhood skill acquisition is an important development (see, for example, Cunha and Heckman, 2007; Bernal and Keane, 2005; and Todd and Wolpin, 2007).

rates of children in the treatment and control villages one year after the program, ranged from 5 to 15 percentage points depending on age and sex (Behrman et al., 2006).

An important limitation of large scale social experiments, such as PROGRESSA, is that it is often prohibitively costly to vary the experimental treatments in a way that permits evaluation of a variety of policies of interest. For example, in the case of PROGRESSA, the policy maker presumably would like to know how program impacts would vary with changes in the level of subsidies or with a change in the performance criterion, for example, by providing subsidies for grade completion rather than school attendance. In the PROGRESSA experiment, all eligible treatment group households faced the same subsidy schedule, so it is not possible to evaluate the effects of such changes through simple comparisons of treatments and controls. In addition, because the experiment lasted only two years, one cannot directly assess the long term impacts of the program on completed schooling. A major goal of the Todd and Wolpin paper was to provide quantitative estimates of the impact of the program along these additional dimensions. To this end, they estimated a DCDP model of parental decision-making about children's schooling and fertility and used the model to compare the effects of the existing subsidy program to the effects of various alternative (non-existent) programs that were simulated from the model.

In their model, each year a married couple decides on whether each child between the ages of 6 and 15 will attend school, remain at home or, for those aged 12–15, work in the labor market. These are assumed to be mutually exclusive activities. They also decide whether the wife will become pregnant (while fecund). The model allows for up to 8 children. The couple receives utility in each period from their current stock of children, their children's current years of schooling, their current school attendance interacted with the number of years they've completed and whether they are behind in school for their age, and from the set of children at home by their ages and gender. There is also a utility cost to attending school (grades 7–9) that depends on the distance from the village to a school. Households differ in their preferences for the choice variables according to their discrete "type" and household preferences are subject to mutually serially independent time-varying shocks.

The household's income is composed of the parent's income, which is assumed to be a function of the age of the household head, the distance of the village to the nearest city, the household's type and a serially independent shock, and the wage income of the children who work, where a child's wage (offer) depends on the child's age and sex, the distance to the nearest city, household type and a serially independent shock. Not all children who attend school complete the grade and the PROGRESSA program only pays for attendance in the same grade with one repetition. The model, therefore, includes a failure probability function, that depends on the child's grade level, age and the household type.

A baseline survey was conducted in October 1997 (with a follow-up baseline survey in March 1998) of all households in both the treatment and control villages prior to the implementation of the program. The program began in the 1998/1999 school year and the experiment continued for that academic year and the next. During the experiment, surveys were conducted in October 1998, March 1999 and November 1999. Treatment households were not informed about the program until after the baseline surveys were completed and control families were not informed about the existence of the program. Within the treatment villages, only households that satisfied an eligibility criterion based on a "marginality" index were provided with the subsidy. The program (which included a child health component as well) provided benefits that, on average, amounted to about 25% of family income.[23] The school attendance subsidy schedule is shown in Table 4. The subsidy begins at grade 3 and rises with each additional completed year of schooling to offset the increased opportunity cost of attending school as children become older. The subsidy level is the same for girls and boys up to grade 6, but then is larger for girls in grades 7–9, reflecting the greater tendency for girls to stop at grade 6.

Todd and Wolpin used data on the treatment households before the program was initiated and the control households both before and after.[24] Thus, none of the families used in the estimation were subject to the subsidy, or knew about the program, at the time they made the decisions about their children's time allocation or their own fertility. As we discuss below, data on the behaviors of the treatment households after the implementation of the program are used to validate the DCDP decision model. If data on the treatment households after the subsidy was introduced had been used in estimation,

Table 4
Monthly transfers for school attendance under the PROGRESSA program

School level	Grade	Monthly payment in pesos	
		Females	Males
Primary	3	70	70
	4	80	80
	5	105	105
	6	135	135
Secondary	1	210	200
	2	235	210
	3	255	225

Corresponds to the first term of the 1998–1999 school year.

[23] The school attendance subsidy component amounted to about 75% of total payments.
[24] For reasons discussed in Todd and Wolpin, the sample was also restricted to landless households.

Table 5
The effectiveness and cost of alternative education policies

	Baseline	Compulsory attendance	PROGRESSA subsidy	Modified subsidy	Bonus for completing 9th grade	Build schools	Subsidy + 25% increase in child wage
Mean completed schooling							
Girls	6.29	8.37	6.83	6.97	6.50	6.39	6.75
Boys	6.42	8.29	6.96	7.07	6.58	6.55	6.79
Pct. completed 9th grade							
Girls	19.8	55.5	25.9	29.3	28.8	21.2	25.2
Boys	22.8	54.7	28.0	31.8	32.7	24.1	26.6
Cost per family	0	?	26,096	25,193	36,976	?	25,250

the model would have been identified from variation in the direct cost of schooling, zero before the program and the subsidy (negative cost) after the program. Given the estimation sample, however, for which there is no direct variation in the cost of schooling, model identification relies on the arguments provided in Section 2, namely on there being some variable(s) that affect the child wage (offer), but that do not affect parental preferences. Todd and Wolpin use the distance of the village from a city as one of the identifying variables.[25]

Table 5 compares the predicted effects on completed schooling of the PROGRESSA program, as implemented, with alternative programs. As the first column shows, the model predicts that the average years of completed schooling in the absence of the program would be 6.29 for girls and 6.42 for boys and that 19.8% of girls and 22.8% of boys would have completed the 9th grade.[26] The first counterfactual (column 2), the perfect enforcement of a compulsory school attendance law, establishes a maximal program impact. As seen, given predicted failure rates (which are lower for girls), average completed schooling would be at most 8.37 years for girls and 8.29 for boys. The next column shows the predicted effects of the PROGRESSA subsidy schedule. The model predicts an increase in completed schooling of about one-half year for both boys and girls, or 26.0% of the maximal potential increase for girls and 28.9% for boys.[27] The last row of the table reports the per-family government budgetary cost of the program over the lifetime of the families in the sample, that is from the woman's age at marriage to age 59. The model predicts that cost to be 26,000 (1997) pesos.

As noted, the PROGRESSA subsidy schedule rewards school attendance starting at grade 3. However, attendance at ages in which children would normally be attending grades 3–5 is almost universal; thus, this aspect of the program is essentially a pure income transfer. Todd and Wolpin calculated that the per-family cost of the program could be held constant if the subsidy in grades 3–5 were eliminated and the subsidy in grades 6–9 were increased by about 45%. Column four shows that the cost of the grades 3–5 subsidy in terms of foregone completed schooling is 0.14 years for girls and 0.11 years for boys. Moreover, under the modified plan, the proportion of girls completing ninth grade increases by 3.4 percentage points and proportion of boys by 3.8%.[28]

An alternative program, shown in the next column, is to provide a large graduation bonus for completing the 9th grade. Providing such a bonus of 3000 pesos has a smaller impact on average completed schooling than the original subsidy, by 0.33 years for girls and 0.38 years for boys, but a larger impact on the 9th grade completion rate, by 2.9 percentage points for girls and 4.7 percentage points for boys. This additional gain in the proportion of ninth-grade graduates is obtained at a per-family cost that is about 40% more than the original subsidy.

Only 26% of all villages are reported to have junior secondary schools. As noted, in the model, families incur a cost of having children attend grades 7–9 that depends on distance. The benefit, in terms of increased schooling, of building a junior secondary school in all villages that do not have one is shown in Table 5 to be relatively small. For girls, average completed schooling would increase by 0.10 years and the proportion completing 9th grade by 1.4 percentage points; comparable figures for boys are 0.13 years and 1.3 percentage points. Comparing the efficacy of such a program with the other programs presented in the table would require knowledge of school construction costs and depreciation rates.[29]

Finally, as noted in Todd and Wolpin, the analysis does not account for changes in the child labor market that would occur as the incentive to attend school increases and children withdraw from the labor market. The last column provides some evidence on the countervailing impact that an increase in the child wage would have on school completion levels. Simulating the PROGRESSA program allowing for an increase in the child wage of 25% would reduce the increase in average

[25] The model forecasted the subsidy effect accurately for girls, but overstated the effect for boys.

[26] These figures are based on simulations of households from the time of marriage until the last born child reaches age 16.

[27] Interestingly, this estimate corresponds closely with that obtained by Behrman et al. (2006) based on grade transition rates for the treatment vs. the control households.

[28] Offsetting this increase in 9th grade completion rates is a small fall in the proportion of children who complete at least 6th grade, by 0.3 percentage points for girls and 0.5 for boys.

[29] Additional experiments were a pure income transfer (found to be extremely cost-ineffective in raising schooling levels), a doubling of the subsidy at all grades and a halving of the subsidy at all grades.

schooling from 0.54 to 0.46 years for girls and from 0.54 to 0.37 years for boys. The larger impact for boys results from their more elastic labor supply response. Whether a 25% increase in the child wage is a reasonable guess is unclear. Jointly modeling the equilibrium in the rural labor market would be a useful, though, non-trivial extension.

In assessing the value of the DCDP structural approach, it is useful to recall the finding about the effect on schooling of a budget-neutral shift in resources towards the higher grades. That such a shift would increase schooling overall must be true from the fact that the attendance up to grade five is essentially universal without the subsidy. However, it is not possible to determine the impact quantitatively from the experiment alone. A policy-maker with limited resources could not make an informed decision about whether to continue the subsidy to the lower grades, given its redistributive function, without knowing its quantitative trade-off with foregone schooling. The structural approach permits a quantitative cost-benefit analysis of alternative programs without having to implement them.

3.4. ***Question:*** *How will the new expansion of Medicare to incorporate prescription drug coverage affect the health and life expectancy of the elderly?*
Claudio Lucarelli (2006)

In 2006, the US introduced the largest expansion in the history of Medicare. The Medicare Prescription Drug Improvement and Modernization Act (MMA) expanded medical coverage for the elderly to include prescription drugs, at an expected cost of $700 billion dollars over ten years. In enacting this legislation, proponents argued that providing prescription drug benefits would not only improve the health of the elderly, but would reduce the use of substitute medical services in the form of outpatient and inpatient care.

The details of the plan are provided, as in Lucarelli, in Fig. 1. Each senior with an income of above 150% of the poverty line pays an annual premium of 420 dollars and also pays the first 250 dollars of prescription drug costs out of pocket. Medicare pays 75% of the of the next 2000 dollars, nothing for the next 2850 dollars (the so-called donut hole), and 95% of all drug costs above 5100 dollars (3600 dollars out of pocket). To induce seniors to sign up for the plan early, the premium increased 1% per month after a fixed grace period of several months.

Lucarelli develops and estimates a DCDP model, relevant prior to the introduction of the prescription drug benefit, in which the elderly choose among seven alternative health insurance options as well as expenditures on prescription drugs and outpatient care. Health insurance options vary by their premium and coverage. The per-period utility flow depends on a subset of the health insurance choices, for example, being covered by an HMO which restricts the range of health providers, on whether there is a switch in health insurance from the previous period (representing the cost of learning about a new plan), on the individual's health status and on consumption. There are permanent differences in preferences represented by distinct types and time varying preference shocks.[30] Income, which is determined by age, education and a time-varying stochastic shock, is spent on the health insurance premium for the chosen plan, on out-of-pocket expenditures on health care (prescription drugs, outpatient care and inpatient care), which depend on the chosen health insurance option, and on consumption goods. Health status transitions (including transitions to death) are subject to a health production technology that includes as inputs expenditures on drugs, outpatient care and inpatient care and that also depends on whether the individual smoked in the past and on age. The individual solves the optimization problem to a fixed terminal age, 100, subject to the mortality risk generated by the health transition production function.

The data used for the analysis are from the Medicare Current Beneficiary Survey for the years 1994–1999, a rotating panel of 12,000 Medicare beneficiaries. Respondents are observed for three years within any given panel. The sample consists of

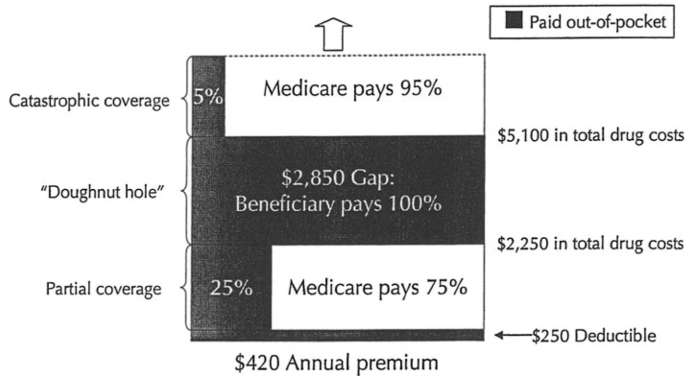

Fig. 1. Medicare part D coverage for seniors with income above 150% of poverty line.

[30] Unlike in the other models, the preference shocks are attached to mutually exclusive choices rather than to the individual dichotomous choices. Lucarelli assumes these shocks to be distributed extreme value, in which case, as Rust (1987) first demonstrated, the solution to the DCDP problem and the likelihood function have closed forms.

Table 6
The effect of the Medicare expansion on medical expenditures and health status

	All	Survivors before policy change	Non-survivors before policy change
Mean expenditures on prescription drugs			
Before	950	990	690
After	1180	1120	1450
Outpatient care			
Before	2680	2830	1890
After	2770	2710	3200
Inpatient care			
Before	2530	2770	1590
After	2540	2390	3150
Proportion in bad health			
Before	0.325	0.304	0.470
After	0.332	0.287	0.630[a]

[a] During extra years of life.

1800 women age 68 and over who are living alone; their average age is 79. Within the sample, 44% of the women have drug prescription coverage through their health insurance plan. Estimation is by maximum likelihood.

Given parameter estimates, the decision model is simulated with the new Medicare drug plan in place. The simulation exercise compares a 10-year forecast of the outcomes of the decision model with and without the drug plan. Table 6 summarizes some of the results. As seen in the first column, expenditures on drugs would rise, on average, by 24%, from 950 to 1180 dollars per year. However, seemingly contrary to the expectations of the proponents of the plan, expenditures on outpatient medical care would also rise, by 3%, while that on inpatient care would remain essentially unchanged. Moreover, the failure of these other medical expenditures to fall occurs even though prescription drugs are found to be a substitute for outpatient and inpatient expenditures in producing health. The reason, as Lucarelli demonstrates, is compositional, namely that while the plan actually succeeds in prolonging life, those who otherwise would have died earlier spend more on medical care.

Table 6 contrasts the before and after expenditures for those simulations in which a person's life is not prolonged after the introduction of the plan (column 2) to those simulations in which the persons' life is prolonged (column 3). The results are striking. In the simulations in which lives were not prolonged, annual average expenditures on drugs (over the ten year period) increase by 13%, and, as expected, expenditures on outpatient and inpatient care fall, by 4 and 14% respectively. But, in the cases in which life is prolonged, expenditures on drugs increase, by 110%, as do expenditures on outpatient care, by 69%, and on inpatient care, by 98%. Moreover, as a result of the increased life span, the proportion of periods spent in poor health actually increases from 0.470 to 0.630. This contrasts with a slight reduction in the proportion of periods spent in bad health from 0.304 to 0.289 for those simulations in which life span did not improve. Thus, while the drug plan extends lifetimes, it does so more for those who would have been in bad health without the program and does so without improving health status in those additional years of life. Lucarelli calculates that the program cost of an additional year of life is 38,000 dollars.

It is unclear as to how, if at all, the government conducted a prospective cost-benefit analysis of the Medicare expansion. A thoughtful consideration of the program would have recognized the existence of offsetting effects of the program on the use of medical services due to the effect of prescription drugs on the survival of those in poor health. It is doubtful that the government had access to quantitative estimates of the extent of the offset as provided by Lucarelli's analysis.

3.5. ***Question:*** *How would the career decisions of politicians be influenced by such policies as term limits, changes in Congressional salaries, and changes in post-Congressional career options?*

Daniel Diermeier et al., (2005)

Political scientists have long been interested in the factors that motivate politicians. For example, a member of the US Congress may be interested in holding office for its own sake (i.e., prestige and perks of office), may care about influencing policy and achieving legislative accomplishments, and/or may care about the monetary rewards that come from serving in Congress. The latter include not only the Congressional salary itself, but also Congressional pension benefits, and the possibility that Congressional experience may be valuable in post-Congressional employment (for example, as a lobbyist). The relative weight that politicians place on these different types of rewards is important, because it influences how they might react to policy changes.

If, for example, there are different "types" of politicians, who weight these rewards differently, then a pay raise designed to encourage "good" politicians to stay in Congress might instead only encourage those who care most about monetary rewards, while having little impact on those who care about policy and legislative accomplishments. Or, if legislative accomplishments take a long time to attain, then term limits may disproportionately induce those who most value such accomplishments to leave Congress. Thus, a failure to understand the motivations of politicians may lead to misguided policies.

In order to gain a better understanding of the motivations of members of Congress, Diermeier et al. (2005) develop a model of the career decisions of members of Congress that includes all of the potential motivating factors described above. Their model takes it as given that a politician is already a member of Congress—i.e., they do not attempt to model the initial decision to run for Congress, but only the career decisions of sitting members. The decision period in the model is two years, the length of a term in the US House. At the end of each term, a House member decides whether to run for re-election, exit Congress or, if the option is available, run for higher office (the Senate). If the politician exits Congress, there are three post-Congressional career options: a private sector job (e.g., lobbyist, lawyer), a political job (e.g., a judgeship, a cabinet post) or complete retirement.

A member of the Senate faces a different choice set. The Senate term is 6 years. At the end of each 6 year term, Senators decide whether to run for re-election or exit (Senators can also choose to exit 2 or 4 years into their term). Upon exiting, a Senator faces the same three post-Congressional career options as does a member of the House.

Members of Congress receive a flow utility in each two-year period that consists of the Congressional wage plus the monetized value of the non-pecuniary rewards from sitting in the House or Senate (i.e., prestige and perks of office) plus their utility from any legislative accomplishments. They also receive utility/disutility from running for office. Upon exiting Congress, former members receive a flow utility that depends on which option they choose. If they choose either the private or public sector occupation, they receive a draw from an occupation specific earnings distribution. The mean of each distribution depends on Congressional experience, as well as on other characteristics of the politician (e.g., education, age). If the politician chooses retirement, then he/she receives the Congressional pension (which again depends on Congressional experience) along with the utility from leisure. Diermeier et al. (2005) calculate a present value of rewards from the date of retirement until death for each of the three options.

Diermeier et al. (2005) estimate their model of career decisions of politicians using data on everyone who entered the US Congress from 1947 (the 80th Congress) to 1993 (the 103rd Congress), giving 1899 career histories. Three aspects of the data collection are important. First, Congressional wages and pension rules varied over the sample period. These data are readily available, enabling Diermeier et al. to identify how these factors influenced decisions of members of Congress. Second, using data from several sources, Diermeier et al. construct, for a subset of the sample, information on post-Congressional career choices and salaries. This enables them to estimate how post-Congressional wages and occupational choices are influenced by Congressional experience. The post-Congressional wage equations are estimated jointly with the rest of the model to deal with the selection bias that arises because (i) leaving Congress is either a choice or results from electoral defeat, and in neither case is it an exogenous random process, and (ii) the post-Congressional employment option is also a choice.

Third, Diermeier et al. (2005) use a monumental data set constructed by David Mayhew (2000) on important legislative accomplishments by members of Congress. Using this data, they can estimate both how the probability of an accomplishment depends on a politician's characteristics (e.g., it is increasing in seniority), and the utility that politicians get from accomplishments. Thus, subject to the assumptions of their model, Diermeier et al. are able to parcel out the utility flows that members of Congress get from (i) wages, (ii) the impact of Congressional experience on post-Congressional wages and pensions, (iii) legislative accomplishments and (iv) sitting in Congress (i.e., perks and prestige).

The politicians in the Diermeier et al. (2005) model make their decisions subject to three important constraints. In addition to the pension rules and post-Congressional wage functions mentioned earlier, the third is the probability of winning elections. This depends on the politician's type ("skilled" or "unskilled"), seniority, age, committee membership and the political climate, as well as whether the politician may have been subject to a scandal or redistricting.

Finally, the Diermeier et al. (2005) model allows politicians to differ on two latent dimensions. The first is whether the politician is "skilled" or "unskilled" at winning elections. The second is whether the politician is the type who values personal legislative accomplishments. The latter distinction is motivated by the data, since, according to Mayhew (2000), the vast majority of important legislative accomplishments are realized by a relatively small fraction of politicians. Diermeier et al. estimate that 27% of politicians are the "achiever" type, while 46% are the skilled type.

According to Diermeier et al.'s estimates, the non-pecuniary reward from sitting in the House is $20,000 to $35,000 per year (depending on whether one belongs to an important committee), while that for a Senator is $200,000 per year. The non-pecuniary reward from an important legislative accomplishment is close to $400,000. Recall that only "achievers" can receive these rewards. The probability an achiever realizes such an accomplishment is about 7.5% per term, so the expected reward is roughly $30,000 per term. To put these figures in perspective, note that the average Congressional salary during our sample period was $120,000 (in 1995 dollars). Diermeier et al. (2005) also find that winning re-election in the House (Senate) for the first time increases post-Congressional wages in the private sector by 4.4% (16.7%). However, the marginal effect of Congressional experience on post-Congressional wages diminishes quite rapidly with additional experience; averaging over members' actual experience levels, the marginal effect on post-Congressional wages of an additional term in the House (Senate) is equal to 2.4% (5.2%).

Having estimated their model, Diermeier et al. (2005) use it to predict the impact of several "policy" experiments. Table 7 reports on two of them, the introduction of term limits and a reduction in Congressional salaries. With respect to term limits, they consider imposing a maximum of 4 terms in the House and 2 terms in the Senate. As seen in the table, term limits reduce the average duration of Congressional careers from 9.61 to 6.06 years. Part of this effect is mechanical (i.e., politicians running up against the term limit), but a large part is also behavioral. Forward looking politicians realize that

Table 7
Selected effects of term limits and Congressional salaries

	Baseline	Term limits[a]	Reduce salaries 20%
Years in Congress			
1. All	9.61	6.06	8.27
2. Achiever type	10.56	6.72	9.27
3. Non-achiever type	9.29	5.82	7.93
2÷3	1.14	1.15	1.17
4. Democrat	10.06	6.00	8.36
5. Republican	9.10	6.13	8.17
4÷5	1.11	0.98	1.02
Value of a seat[b]			
House	616,228	418,322	491,905
Senate	1,673,763	1,322,786	1,512,592

[a] 4 terms for House, 2 terms for Senate.
[b] Monetary + non-monetary compensation (1995 dollars).

the value of a seat in Congress has dropped (by 40% and 26% in the House and Senate, respectively, according to Diermeier et al.'s estimates). Thus, there is a substantial increase in voluntary early exit from Congress. Specifically, Diermeier et al. (2005) estimate that the probability a member of the House runs for re-election (for a second or third term) drops from 93.6 to 81.9%. The probability a House member runs for Senate increases from 2.9 to 6.7%. Thus, the probability of voluntary exit from Congress increases from only 3.5 to 11.4%.

The second policy experiment they consider is a 20% reduction in Congressional salaries. The impact on the average duration, although not as large as for term limits, is nevertheless substantial; average duration falls from 9.61 to 8.27 years. This decline results from an increase in the probability of a voluntary exit from the House from 6.7 to 10.3% and from the Senate from 14.8 to 17.7%.

More interesting is how these changes would shift the composition of Congress, that is, the extent to which they would differentially increase the probability of exit for different types of politicians. In a subsequent paper, Keane and Merlo (2007) provide more detail about how different types of politicians are affected by terms limits and Congressional salaries, as well as by a range of other policies not considered by Diermeier et al. (2005). For instance, as seen in Table 7, a reduction in salaries and the imposition of term limits would have similar effects on achievers and non-achievers. Compared to the baseline case, in which the average Congressional career duration of achievers exceeds that of non-achievers by 14%, reductions in salaries or term limits would only very slightly increase the relative career durations of achievers to 17% and 15% respectively. Thus, if we take being an "achiever" as a good measure of the "quality" of a politician, there is no evidence to suggest that term limits would disproportionately induce higher quality politicians to exit Congress.

An especially interesting consequence of these changes is that they lower the average career duration of Democrats by considerably more than of Republicans. Under the baseline, Democrats have average career durations that are 11% greater than Republicans. Reducing Congressional salaries by 20% reduces the relative longevity of Democrats to only 2%. The effect of term limits is even larger; under term limits, the average Congressional duration of Democrats is 2% less than that of Republicans. The effect of term limits is partly because Democrats were the majority party during the whole estimation period. Hence, by construction, they have higher probabilities of winning elections, and tend to have longer careers than Republicans (10 years vs. 9 years). Thus, they are more likely to hit up against the term limits. This illuminates a feature of term limits that seems rather obvious *ex post*, but which, to our knowledge, had not been previously noted. Term limits will tend to favor the minority party relative to the majority party. However, as illustrated by the impact of the reduction in salaries, this is not the entire effect. Democrats are simply more sensitive to reductions in the value of a seat than are Republicans.

As Table 7 also shows, the value of a seat falls significantly under these two policies. Although the 20% reduction in salaries causes about the same size reduction in the value of a House seat, because of the high non-pecuniary value of a Senate seat, the reduction in its value is only 9.6%. Term limits induce an even larger reduction in the value of Congressional seats, 32% for a House seat and 21% for a Senate seat.

3.6. ***Question:*** *What would be the impact of employer sanctions and increased border protection on the flow of illegal Mexican immigrants to the US?*
Aldo Collusi (2006)

Mexican illegal immigration has been, and continues to be, a central focus of US policy concern. From 1942 to 1964, under the Bracero program, the US provided a fixed annual number of temporary visas to Mexican agricultural workers. The number of visas peaked at about 400,000 annually between 1955 and 1960. The program was abandoned in 1965 and Mexican legal migration dropped precipitously. The first major piece of legislation subsequent to abandoning the Bracero program was the Immigration Reform and Control Act of 1986. That legislation introduced employer sanctions for hiring illegal immigrants, increased the level of resources for border control and introduced an amnesty program for illegal immigrants who could demonstrate that they had continuously resided in the US since 1982. A series of legislative and administrative initiatives were introduced throughout the 1990s in an attempt to reduce the flow of illegal immigrants. And, major new pieces of Congressional legislation have recently been proposed. As before, the major tools to reduce illegal immigration being debated in the proposed legislation include border enforcement and employer sanctions.

Colussi (2006) addresses the efficacy of these alternative policies by estimating a labor market equilibrium model of migrant decision-making that embeds a DCDP decision model. The village agricultural economy consists of overlapping generations of 15 to 64 year old potential migrants. A potential migrant chooses between working in the US, subject to receiving a job offer, or remaining in his Mexican village as an agricultural worker. Crossing the border entails a cost and residing in the US may entail a non-pecuniary cost (or benefit). The village economy is modeled as having a closed labor market, so that migrants to the US reduce the supply of labor in their home village and put upward pressure on village wages. Return migration occurs because the utility of residing in the US declines with the current duration in the US as well as from a response to transitory shocks to US wages.

Migrants from any single village have no effect on US wages and take the (stochastic) wage process in the US as given. However, potential migrants do not necessarily receive job offers. The probability that a potential migrant receives a wage offer in the US depends both on his own personal characteristics, such as his US work experience and on his village network. The more migrants from his home village that are currently working in the US, the higher the likelihood of receiving a job offer. The import of this network effect is that, starting from a situation in which there are no (or few) migrants from the village, as was the case in 1965 after the abandonment of the Bracero program, there first will be a period of rapid growth in migration followed by a slowdown until a steady state is reached. This S-shape to the temporal pattern of migration, that is, to the fraction of migrants from a village, would not arise without the existence of a network effect. The model is solved for the equilibrium fraction of migrants and the equilibrium village wage starting from a situation in which there are no migrants until a steady state is reached.

The model is estimated using data from the Mexican Migration Project. In each year since 1987, a number of villages are chosen for inclusion in the project and a randomly selected set of households within each village is surveyed. A migration history is obtained for the household head, along with demographic and wage information. Colussi estimates the model separately for three villages surveyed in 1988. The pattern of migration for those three villages is depicted in Fig. 2. The initiation and later abandonment of the Bracero program is clearly seen, as is the subsequent rapid rise in migration, particularly in village one, and the convergence to a steady state. The vast majority of border crossings were by undocumented migrants; undocumented crossings account for 60% of all reported crossings for village two, 73% for village three and 85% for village one.

Given the estimates of the model, Colussi simulates the effect of increasing enforcement on the flow of illegal immigrants. Table 8 shows the steady-state impact of two policy interventions on the fraction of migrants and on their average trip duration. In the first, the cost of crossing the border is increased from the baseline estimate of one week's US wages to one month's wages and to six month's wages. In the second, migrant wages in the US are reduced by 10% and by 25%, assuming that the lower wages would be an outcome of increased employer sanctions. As the table shows, increasing the cost of a border crossing to the equivalent of one month's US wages would actually increase the fraction of migrants in the steady state by 9%. This increase arises because, at this higher cost per crossing, while it is optimal for migrants to make fewer trips, the trips are of longer duration. The average duration of stay increases from 5 to 7 years, a large enough increase so that the fraction of migrants is actually higher in the steady state. A further increase in cost, to six month's wages, has even a larger impact on trip duration, increasing it to an average of 19 years, but reduces the number of crossings sufficiently to

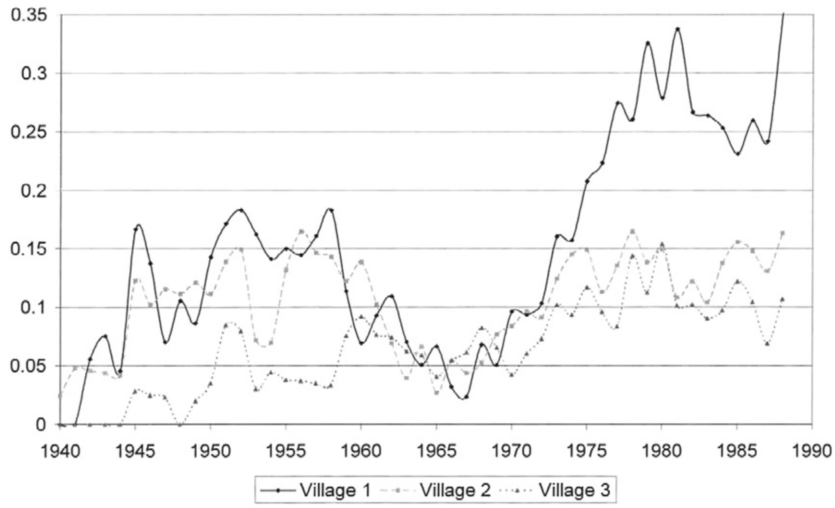

Fig. 2. Percentage of villagers in the US by year.

Table 8
The effect of increased enforcement on illegal immigration (village 1)

	Baseline[a]	Cost of border crossing		US wage reduction	
		One month US wages	Six months US wages	10%	25%
Steady state fraction in US	0.32	0.35	0.28	0.27	0.02
Ave. duration in US (yrs.)	5	7	19	4	2

[a] Cost of border crossing = one week of US wages.

bring about a reduction in the steady-state fraction of migrants by about 12%. Thus, the steady-state fraction of migrants appears to be relatively invariant to the cost of crossing the border.

Reductions in the US wage reduces both the fraction of migrants and their average length of stay in the steady state. A 10% reduction in the wage reduces the fraction of migrants by 16% and the average trip duration by 20%. While these changes are relatively modest, a 25% reduction in the wage almost eliminates migration, with the fraction of migrants declining from 0.32 in the baseline to 0.02. Although a direct comparison of the efficacy of these two policies is not possible without knowing the cost of implementing them, the results indicate that migration flows are more sensitive to wage changes, cum employer sanctions, than to increases in the cost of crossing the border, cum building fences.

4. The credibility of empirical findings from DCDP models

As we have illustrated, applications of the DCDP approach have addressed challenging and important questions 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 a non-parametric assumption-free ideal that does not and is likely never to exist.

Can one provide convincing evidence about the credibility of these exercises? Put differently, how can these 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.[31] That is not to say that evidence accumulated from the estimation of models by different researchers, each with different modeling prejudices, is not valuable.[32] Indeed, 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, based on 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 formal statistical tests. However, it is interesting to note that formal tests generally reject DCDP models nevertheless. Although these models tend to have a lot of parameters, sometimes numbering into the hundreds, given the extensiveness of the data that these models attempt to account for, the models are actually parsimonious; researchers in practice 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. Similarly, Lucarelli estimated his model of the new Medicare drug prescription program, as described above, prior to its introduction and compared the early data on take-up rates to that predicted by his model. Both of these examples exploited data that were unavailable for estimation to validate their model.

[31] Of course, robustness by itself cannot be conclusive; all of the models could give similarly biased results.

[32] Such prejudices are revealed by the contrast between the structure of the DCDP model that Todd and Wolpin (2006), presented above, used to evaluate the Mexican Progressa program and that of Attanasio et al. (2005). As another example, there are several applications of DCDP models applied to traditional topics that take a behavioral economics view. Paserman (2007) study a job search model and Fang and Silverman (2007) study a model of women's welfare participation assuming that agents use hyperbolic discounting.

Some researchers have deliberately held out data to use for validation purposes. Lumsdaine et al. (1992), for example, estimated a model of the retirement behavior of 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 US 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), in the paper discussed above, 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.[33]

As should be clear from this discussion, model validation, and model building more generally, are necessarily part art and part science. Given that, researchers will attach different priors to a model's credibility and different weights to the validation evidence. They will, therefore, come 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, researchers should strive to provide as much validation evidence as the data and methods permit.

5. Conclusion

Our goal in this unconventional survey was not to simply summarize the theory and applications of the DCDP framework. Rather, we set out to provide evidence on whether the methodology has borne fruit as a tool for empirical research. To this end, we discussed six example papers in which DCDP models have been used to provide quantitative evidence on substantively important and challenging economic questions. We think that all six papers provide useful insights. For example, we have seen that, in plausible economic models, (i) a college attendance bonus sufficient to equalize black/white differences in educational attainment would do little to eliminate black/white wage differences; specifically, the gap in lifetime earnings is reduced by only 5%, (ii) the increased supply of women to the labor market induced by changes in home productivity would explain about half of the closing of the male–female wage gap that has occurred over the last 30 years, (iii) altering the schedule of school attendance bonuses in the Mexican Progressa program in a budget neutral way would have increased the impact of the program on completed schooling levels by about 25% for a sample of rural landless households, (iv) because of the increase in life expectancy of those in poor health, the cost of the new Medicare drug benefit would not necessarily be offset by a reduction in Medicare part A and B costs; for a sample of elderly single women, out-patient care costs would actually rise by 3%, (v) imposing Congressional term limits would have wiped out the Democrat majority that existed over (most of) the last half of the 20th century and (vi) increased spending on border enforcement would not necessarily reduce the steady-state level of illegal immigrants in the US; at least for the residents of one village in Mexico, in response to an increase in the cost of a border crossing from 1 week to 1 month of US wages, the average length of stay in the US would increase from 5 to 7 years. Researchers will of course differ as to the credibility of these findings. Regardless, we believe these papers should inform and challenge our thinking about the phenomena they study and the policies they evaluate. And, hopefully, they will serve to stimulate further research.

On a methodological note, as these papers demonstrate, our ability to estimate DCDP models has greatly expanded since the original development. Nevertheless, the research agenda associated with the development of new methods for the estimation of DCDP models continues to be an active one.[34] Methods for the estimation of dynamic equilibrium models and models with strategic interactions in a dynamic setting are still in formative stages, and there needs also to be additional methodological work on model validation/selection. We are optimistic that the next 25 years will see advances in structural estimation even greater than those we have experienced over the past 25 years.

[33] Similarly, Lise et al. (2005) 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.

[34] See, for example, the recent work by Imai et al. (2007) and Norets (2008) for Bayesian approaches to the estimation of DCDP models and by Arcidiacono and Miller (2008) for a recent extension of the Hotz and Miller (1993) "conditional choice probability" methodology.

References

Aguirregabiria, V., Mira, P., in press. Dynamic discrete choice structural models: A survey. Journal of Econometrics.
Arcidiacono, P., Miller, R.A., 2008. CCP estimation of dynamic discrete choice models with unobserved heterogeneity. Mimeo. Duke University.
Attanasio, O., Meghir, C., Santiago, A., 2005. Education choices in Mexico: Using a structural model and a randomized experiment to evaluate Progressa. Mimeo. University College London, April.
Bajari, P., Hortacsu, A., 2005. Are structural estimates of auction models reasonable? Evidence from experimental data. Journal of Political Economy 113, 703–741. August.
Behrman, J., Sengupta, P., Todd, P., 2006. Progressing through PROGRESSA: An impact assessment of a school subsidy experiment in Mexico. Economic Development and Cultural Change 54, 237–276. October.
Ben-Porath, Y., 1967. The production of human capital and the life cycle of earnings. Journal of Political Economy 75, 352–365. August.
Bernal, R., Keane, M.P., 2005. Child care choices and children's cognitive achievement: The case of single mothers. Mimeo. Northwestern University.
Colussi, A., 2006. Migrants' networks: An estimable model of illegal Mexican migration. Mimeo. University of Western Ontario, January.
Cunha, F., Heckman, J.J., 2007. Formulating, identifying and estimating the technology of cognitive and non-cognitive skill formation. American Economic Review 97, 31–47. May.
Diermeier, D., Keane, M.P., Merlo, A., 2005. A political economy model of Congressional careers. American Economic Review 95, 347–373. March.
Eckstein, Z., Wolpin, K.I., 1989a. The specification and estimation of dynamic stochastic discrete choice models. Journal of Human Resources 24, 562–598. Fall 1.
Eckstein, Z., Wolpin, K.I., 1989b. Dynamic labor force participation of married women and endogenous wage growth. Review of Economic Studies 56, 375–390. January.
Fang H., Silverman, D., 2007. Time inconsistency and welfare program participation: Evidence from the NLSY. Working paper W13375. NBER. September.
Gotz, G.A., McCall, J.J., 1984. Estimating the stay/leave decision in a sequential decision-making model. RAND Corporation, Santa Monica, CA.
Heckman, J.J., 1979. Sample selection bias as a specification error. Econometrica 47, 153–162. January.
Heckman, J.J., 1981. The incidental parameters' problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process and some Monte Carlo evidence. In: Manski, C.F., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA, pp. 179–197.
Heckman, J.J., Lochner, L., Taber, C., 1998. Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of earnings with heterogeneous agents. Review of Economic Dynamics 1, 1–58. January.
Hotz, J., Miller, R.A., 1993. Conditional choice probabilities and the estimation of dynamic models. Review of Economic Studies 60, 497–530. July.
Imai, S., Jain, N., Ching, A., 2007. Bayesian estimation of discrete choice models. Mimeo. Queens University.
Katz, L.F., Autor, D.H., 1999. Changes in the EAGE structure and earnings inequality. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, vol. 3. North-Holland, Amsterdam, pp. 1463–1555.
Keane, M.P., Merlo, A., 2007. Money, political ambition and the career decisions of politicians. Mimeo. University of Pennsylvania, May.
Keane, M.P., Moffitt, R., 1998. A structural model of multiple welfare program participation and labor supply. International Economic Review 39, 553–590.
Keane, M.P., Wolpin, K.I., 1997. The career decisions of young men. Journal of Political Economy 105, 473–522. June.
Keane, M.P., Wolpin, K.I., 2000. Eliminating race differences in school attainment and labor market success. Journal of Labor Economics 18 (4), 614–652.
Keane, M.P., Wolpin, K.I., 2001. The effect of parental transfers and borrowing constraints on educational attainment. International Economic Review 42, 1051–1103. November.
Keane, M.P., Wolpin, K.I., 2002. Estimating welfare effects consistent with forward looking behavior. An article in two parts. Journal of Human Resources 37, 570–622. Summer.
Keane, M.P., Wolpin, K.I., 2007. Exploring the usefulness of a non-random holdout sample for model validation: Welfare effects on female behavior. International Economic Review 48, 1351–1378. November.
Lee, D., 2005. An estimable dynamic general equilibrium model of school, work and occupational choice. International Economic Review 46, 1–34. November.
Lee, D., Wolpin, K.I., 2006. Intersectoral labor mobility and the growth of the service sector. Econometrica 74, 1–46. January.
Lee, D., Wolpin, K.I., in press. Accounting for wage and employment changes in the US from 1968–2000: A dynamic model of labor market equilibrium. Journal of Econometrics.
Leibowitz, A., 1974. Home investments in children. Journal of Political Economy 82, S111–S131. March/April, Part II.
Leslie, L.L., Brinkman, P.T., 1987. Student price response in higher education: The student demand studies. Journal of Higher Education 55 (3), 181–204. March/April.
Lise, J., Smith, J., Seitz, S., 2005. Equilibrium policy experiments and the evaluation of social programs. Mimeo. Queens University.
Lucarelli, C., 2006. An analysis of the Medicare prescription drug benefit. PARC Working Paper Series. University of Pennsylvania. September.
Lumsdaine, R., Stock, J., Wise, D., 1992. Three models of retirement: Computational complexity vs. predictive validity. In: Wise, D. (Ed.), Topics in the Economics of Aging. Univ. of Chicago Press, Chicago.
Marschak, J., 1953. Economic measurements for policy and prediction. In: Hood, W.C., Koopmans, T.C. (Eds.), Studies in Econometric Method. In: Cowles Commission for Research in Economics, vol. 14. Wiley, New York.
Mayhew, D.R., 2000. America's Congress: Actions in the Public Sphere—James Madison through Newt Gingrich. Yale Univ. Press, New Haven.
McFadden, D., 1977. Urban Travel Demand Forecasting Project Final Report, Vol. 5. Institute of Transportation Studies, University of California, Berkeley.
Miller, R.A., 1984. Job matching and occupational choice. Journal of Political Economy 92, 1086–1120. December.
Miller, R.A., 1997. Estimating models of dynamic optimization with microeconomic data. In: Pesaran, M.H., Schmidt, P. (Eds.), Handbook of Applied Econometrics: Microeconomics. Basil Blackwell.
Mincer, J., 1962. Labor force participation of married women: A study of labor supply. In: Aspects of Labor Economics. Princeton Univ. Press, Princeton, NJ, pp. 63–97.
Norets, A., 2008. Bayesian inference in dynamic discrete choice models. Mimeo. Princeton University.
Pakes, A., 1986. Patents as options: Some estimates of the value of holding European patent stocks. Econometrica 54, 755–784. July.
Paserman, M.D., 2007. Job search and hyperbolic discounting: Structural estimation and policy evaluation. Mimeo. The Hebrew University, July.
Rust, J., 1987. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. Econometrica 55, 999–1034. September.
Rust, J., 1993. Structural estimation of Markov decision models. Chapter 51 in: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. North-Holland, Amsterdam.
Rust, J., 1994. Numerical dynamic programming in economics. Chapter 14 in: Amman, H.M., Kendrick, D.A., Rust, J. (Eds.), Handbook of Computational Economics. North-Holland, Amsterdam.
Todd, P., Wolpin, K.I., 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. January.
Todd, P., Wolpin, K.I., 2007. The production of cognitive achievement in children: Home, school and racial test score gaps. Journal of Human Capital 1, 91–136. Winter.

Welch, F., 2000. Growth in women's relative wages and inequality among men: One phenomenon or two. American Economic Review 90, 444–449. May.
Wise, D., 1985. A behavioral model versus experimentation: The effects of housing subsidies on rent. In: Brucker, P., Pauly, R. (Eds.), Methods of Operations Research, vol. 50. Verlag Anton Hain, pp. 441–489.
Wolpin, K.I., 1984. A dynamic stochastic model of fertility and child mortality. Journal of Political Economy 92, 852–874. October.
Wolpin, K.I., 1987. Estimating a structural search model: The transition from school to work. Econometrica 55, 801–817. July.
Wolpin, K.I., 2003. Wage equations and education policy. In: Dewatripont, M., Hansen, L.P., Turnovsky, S.J. (Eds.), Advances in Economics and Econometrics. Cambridge Univ. Press, Cambridge.