Structural Estimation and Policy Evaluation in Developing Countries

Petra E. Todd and Kenneth I. Wolpin

Department of Economics, University of Pennsylvania, Philadelphia, Pennsylvania 19104; email: ptodd@ssc.upenn.edu, wolpink@econ.upenn.edu

Annu. Rev. Econ. 2010. 2:21–50

The *Annual Review of Economics* is online at econ.annualreviews.org

This article's doi:
10.1146/annurev.economics.102308.124345

1941-1383/10/0904-0021$20.00

Key Words

development economics, dynamic discrete choice models, schooling, migration

Abstract

This review discusses the use of discrete choice dynamic programming (DCDP) methods for evaluating policies of particular relevance to developing countries, such as policies to reduce child labor and increase school attendance, improve school quality, affect immigration flows, expand old-age pension benefits, or foster small-business investment through microfinance. We describe the DCDP framework and how it relates to static models, illustrate its application with an example related to conditional cash transfer programs, consider numerous empirical applications from the literature of how the DCDP methodology has been used to address substantively important policy issues, and discuss methods for model validation.

Discrete choice dynamic programming (DCDP) models: models in which agents make choices sequentially over time from a discrete set of alternatives as new information arrives to maximize their expected utility over some time horizon

1. INTRODUCTION

The development of methods for the estimation of discrete choice dynamic programming (DCDP) models that began 25 years ago opened up new frontiers for empirical research in a host of areas, including development economics, labor economics, industrial organization, economic demography, health economics, and political economy. This review shows the value of these tools for empirical policy evaluation in the developing-country context.[1] It describes the DCDP paradigm and its use in previous studies to address diverse, substantively important and challenging policy questions that have particular relevance to developing countries. Although some of these studies assess the impact of existing policies, what could be considered ex post evaluation, they also aim to perform ex ante policy evaluation by assessing the impact of policies outside of historical experience. To the extent that these papers are successful, they provide policy makers with the ability to compare the efficacy of alternative prospective policies.

The questions these studies address are the following (with countries listed in parentheses):

1. What policies are effective in increasing educational attainment and improving school quality? (Mexico, Chile, India)
2. How do government pension programs affect household labor supply? (Indonesia, Chile)
3. What policies are effective in increasing business investments by households? (Thailand, India)
4. How does immigration policy affect the flow of migrants between developing and developed countries? (Mexico–United States)

We proceed first by briefly describing the development of the DCDP framework, showing that it is a natural extension of the latent variable formulation of static discrete choice modeling. To fix ideas, we begin in Section 2 with a static binary choice example of parental choice between a child's school attendance or labor market participation. We then introduce a dynamic element into the decision problem. 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. The added complexity comes from having to solve a dynamic programming problem. Similarly, the computational issues that arise in the estimation of static multinomial choice models carry over to the dynamic case. Section 3 summarizes the model structures and empirical results of the papers that tackle the above questions. Section 4 concludes and discusses the synergies that can be exploited when the data are generated by a randomized social experiment.

2. THE COMMON EMPIRICAL STRUCTURE OF STATIC AND DYNAMIC DISCRETE CHOICE MODELS

2.1. Latent Variable Framework

The development of the DCDP empirical framework was a straightforward and natural extension of the static discrete choice empirical framework. The common structure they

[1]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 & Wolpin 1989; Rust 1994a,b; Miller 1997), and a recent survey describes the methodological work that has been ongoing since that time (Aguirregebiria & Mira 2010).

share is based on the latent variable specification, the building block for all economic models of discrete choice. To illustrate the general features of the latent variable specification, consider a binary choice model in which an economic agent with imperfect foresight, denoted by i, makes a choice at each discrete period t, from t = 1 to T, between two alternatives $d_{it} \in \{0, 1\}$. In the developing-country context, some examples are the choice of whether to have one's child work rather than attend school, whether to work in the formal or informal sector of the economy, or whether to become a guest worker in a developed country or perhaps enter a country illegally. 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 $v^*_{it} \geq 0$ and zero otherwise.

Structural estimation: the recovery of fundamental parameters of behavioral models, such as utility or technology parameters

In its most general form, the latent variable may be a function of three types of variables: $\tilde{D}_{it}$ is a vector of the history of all past choices ($d_{i\tau}$: $\tau = 1, \ldots, t-1$), $\tilde{X}_{it}$ is a vector of contemporaneous and lagged values of J additional variables ($X_{ij\tau}$: $j = 1, \ldots, J$; $\tau = 1, \ldots, t$) that enter the decision problem, and $\tilde{\varepsilon}_{i\tau}$ ($\varepsilon_{i\tau} : \tau = 1, \ldots, t$) is a vector of contemporaneous and lagged unobservables that also enter the decision problem.[2] The agent's decision rule at each age is given by

$$\begin{aligned} d_{it} &= 1 \text{ if } v^*_{it}(\tilde{D}_{it}, \tilde{X}_{it}, \tilde{\varepsilon}_{it}) \geq 0, \\ &= 0 \text{ if } v^*_{it}(\tilde{D}_{it}, \tilde{X}_{it}, \tilde{\varepsilon}_{it}) < 0. \end{aligned} \tag{1}$$

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 is to estimate the (parameters of the) v^*_{it} function, where v^*_{it} itself is not observed.[3] 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 & Wolpin 2007), with the class of models often left unspecified. 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 so-called structural approach specifies and solves an explicit constrained optimization problem, as described below. Parameters are identified with specific structural relationships that led to the latent variable function. DCDP models fall within this second approach. Their introduction is associated with independent contributions by Gotz & McCall (1984), Miller (1984), Pakes (1986), Rust (1987), and Wolpin (1984, 1987). The basic insight of the structural estimation approach is that any DCDP model can be cast as a static estimation problem. To illustrate that insight, it is useful to consider an example.

2.2. Example: Child Labor and Schooling

To demonstrate the connection between the structures of static and dynamic discrete choice models, we consider a simplified example in which a household decides whether to

[2]As 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.

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

send a child to school or to have the child work in the labor market. We first consider a static model, then add a dynamic feature, and show how each falls within the latent variable framework. We then consider several extensions of the dynamic model as well as the value of the model for ex ante policy evaluation.

2.2.1. Static model. Consider a family, i, choosing at time t between sending their only child to work, d_{it} = 1, or to school, d_{it} = 0. 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. The utility the couple attaches to the child's school attendance, α_{it}, is time-varying. Specifically,

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

where the variables in x_{it} ($\subseteq X_{it}$) include, perhaps among other things, the parents' schooling or the child's sex, and where ε_{it} is a serially uncorrelated random shock to the utility of the child's school attendance. The child receives a wage offer of w_{it} in each period t, and the household otherwise generates income y_{it}.

The household budget constraint is

$$C_{it} = y_{it} + w_{it}d_{it}, \tag{4}$$

where, for convenience, we assume that there are no costs associated with attending school (e.g., books, tuition). By substituting Equations 3 and 4 into Equation 2, we can write alternative-specific utilities, U^1_{it} if the child works and U^0_{it} if the child attends school, as

$$U^1_{it} = y_{it} + w_{it},\ \ U^0_{it} = y_{it} + x_{it}\beta + \varepsilon_{it}. \tag{5}$$

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

$$v^*_{it}(x_{it}, w_{it}, \varepsilon_{it}) = w_{it} - x_{it}\beta - \varepsilon_{it}. \tag{6}$$

It is useful to distinguish the household's state space, Ω_{it}, consisting of all the determinants of the household's decision, that is, w_{it}, x_{it}, and ε_{it}, from the part of the state space observable to the researcher, Ω^-_{it}, consisting only of w_{it} and x_{it}. Given Equation 6, the value of ε_{it} at which the couple is indifferent between the child working or attending school is $\varepsilon^*_{it}(\Omega^-_{it}) = w_{it} - x_{it}\beta$. Households with a value of ε_{it} below $\varepsilon^*_{it}(\Omega^-_{it})$ choose to have the child work, whereas households with a value of ε_{it} above $\varepsilon^*_{it}(\Omega^-_{it})$ 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

$$\Pr(d_{it} = 1|\Omega^-_{it}) = \int_{-\infty}^{\varepsilon^*_{it}} dF_\varepsilon(\varepsilon_{it}|\Omega^-_{it}) = \int_{-\infty}^{\varepsilon^*_{it}} dF_\varepsilon(\varepsilon_{it}), \tag{7}$$

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

$$L(\theta x_{it}, w_{it}) = \Pi^{i=I}_{i=1}\Pr(d_{it} = 1|\Omega^-_{it})^{d_{it}}\Pr(d_{it} = 0|\Omega^-_{it})^{1-d_{it}}, \tag{8}$$

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

It would be unusual to have data on the child 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 to proceed with estimation. Let wage offers be generated by

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

where z_{it} ($\subseteq Z_{it}$) would contain, α depending perhaps on the type of labor, the child's age, gender, and physical attributes, or some factor that affects the demand for child labor such as distance to a city, and where the independent and identically distributed (i.i.d.) wage shock has distribution function F_η. We explicitly do not include the child's current educational attainment in z in order to maintain the static nature of the model. We introduce dynamic considerations below.

Substituting Equation 9 into Equation 6 yields the latent variable function

$$v^*_{it}(x_{it}, z_{it}, \varepsilon_{it}, \eta_{it}) = z_{it}\gamma - x_{it}\beta + \eta_{it} - \varepsilon_{it} = \xi^*_{it}(\Omega^-_{it}) + \xi_{it}, \tag{10}$$

where $\xi_{it} = \eta_{it} - \varepsilon_{it}$, $\xi^*_{it}(\Omega^-_{it}) = z_{it}\gamma - x_{it}\beta$, and Ω^-_{it} now consists of z_{it} and x_{it}. 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 semilog form or if the utility function is nonlinear in consumption (e.g., CRRA). The likelihood function incorporates the wage information, namely

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

where $\Pr(d_{it} = 1, w_{it}|\Omega^-_{it}) = \Pr(\xi_{it} \geq -\xi^*_{it}(\Omega^-_{it}), \eta_{it} = w_{it} - z_{it}\gamma)$ and $\Pr(d_{it} = 0|\Omega^-_{it}) = \Pr(\xi_{it} < -\xi^*_{it}(\Omega^-_{it}))$.[4]

It is not instructive, for our purpose, to consider identification of this model in a semiparametric setting, that is, without making distributional assumptions about the underlying unobservables (ε_{it} and η_{it}). To proceed, then, assume that $f(\varepsilon, \eta)$ is joint normal with variance-covariance matrix, $\Lambda = \begin{pmatrix} \sigma^2_\varepsilon & \cdot \\ \sigma_{\varepsilon\eta} & \sigma^2_\eta \end{pmatrix}$. The parameters to be estimated include β, γ, π, σ^2_ε, σ^2_η, and $\sigma_{\varepsilon\eta}$. As is well-known, joint normality is sufficient to identify the wage parameters (γ and σ^2_η) as well as $(\sigma^2_\eta - \sigma_{\varepsilon\eta})/\sigma_\xi$ (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 utility 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 $\sigma_{\varepsilon\eta}$) requires the existence of at least one variable that appears only in the wage equation. In the example, variables such as those that affect the demand for labor must not affect the utility value the couple places on the child's school attendance.

Why should we impose this additional restriction? We can, after all, estimate how school attendance 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 boys versus girls results from differences in parental preferences and how much results

[4]We call $\Pr(d_{it} = 1, w_{it}|\Omega^-_{it})$ a probability, but it is actually a mixed probability for d and a density for w.

from differences in wage 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.

To illustrate this latter point, suppose that the government wanted to implement a policy that would increase school attendance among children. One such policy would be to provide a subsidy to the family if the child attends school. The household budget constraint under the subsidy is

$$C_{it} = w_{it}d_{it} + y_{it} + \tau(1 - d_{it}), \tag{12}$$

where τ is the level of the school attendance subsidy. With the subsidy, the probability that the child works is

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

where Φ is the standard normal cumulative. As seen in Equation 13, it is necessary to have identified σ_{ξ} to predict the effect of the policy, that is, the difference in the probability of participation, and thus school attendance, when τ is positive and when τ is zero. Government outlays on the program would be equal to the subsidy amount multiplied by the number of children who attend school (given by Equation 13).

Importantly, the policy effect can be estimated without direct policy variation; i.e., we did not need to observe families in both states of the world, with and without the subsidy program. Moreover, tuition or other direct costs of attendance, from which we could extrapolate the impact of a subsidy, were assumed to be zero. What was critical for identification was (exogenous) variation in the wage (independent of preferences). In the model, the cost of attending school when there is no subsidy is the foregone wage, whereas when there is a subsidy it is the foregone wage minus the subsidy. The effect of the wage on the decision is isomorphic to the effect of the subsidy; it provides policy-relevant variation.[5]

2.2.2. 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. For example, suppose the child's wage increases with actual work experience, h. In that case, rewrite Equation 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 future wages. Alternatively, or additionally, we could suppose that the child wage and/or parents' utility depends not only on current attendance, but also on the current number of years of schooling the child has completed. In that case, attending school in any period affects future utility either indirectly through the wage or directly.

Continuing with the example, modified to account for the wage offer function in Equation 14 but otherwise maintaining the same structure, assume that the couple maxi-

[5]Todd & Wolpin (2010) show that in this simple setting the policy effect is actually nonparametrically identified. As we discuss below, if households place an intrinsic value (positive or negative) on participating in the program (as in Attanasio et al. 2005), wage variation alone will not be sufficient.

mizes 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.[6] 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,

Bellman equation: a method for solving dynamic optimization problems by breaking them into simpler sequential subproblems

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

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 Equation 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, $V^k_t(\Omega_{it})$, where $k \in \{0, 1\}$, and

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

each of which obeys the Bellman equation

$$\begin{aligned} V^k_t(\Omega_{it}) &= U^k_{it} + \delta E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}, d_{it} = k] \text{ for } t < T, \\ &= U^k_{iT} \text{ for } t = T. \end{aligned} \tag{18}$$

The expectation in Equation 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.[7]

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

$$\begin{aligned} v^*_{it}(\Omega_{it}) = z_{it}\gamma_1 + \gamma_2 h_{it} - x_{it}\beta - \varepsilon_{it} + \eta_{it} + \delta\{[E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}, d_{it} = 1] \\ -[E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}, d_{it} = 0]\} = \xi^*_{it}(\Omega^-_{it}) + \xi_{it}. \end{aligned} \tag{19}$$

Comparing the latent variable function in the dynamic case (Equation 19) with that of the static case (Equation 10), the only difference is the appearance of the difference in the future component of the expected value functions under the two alternatives in Equation 19. A full solution of the dynamic programming problem consists of finding $E[\max(V^0_t(\Omega_{it}), V^1_t(\Omega_{it}))]$ at all values of Ω^-_{it}, denoted by $Emax(\Omega^-_{it})$, for all t = 1,...,T. Given such a solution, as seen in the second equality in Equation 19, the estimation of the dynamic model is in principle no different than the estimation of the static model. The main conceptual difference is that the dynamic problem introduces an additional parameter, the discount factor, δ. Additionally, the dynamic problem requires an assumption on the subjective probability distribution that agents hold for future outcomes under alternative choice paths.[8] The practical difference in terms of implementation may be large.

[6]For example, the terminal period could be when the child reaches adulthood.

[7]We assume for simplicity that individuals have perfect foresight about future values of z_{it} and x_{it}. If z_{it} or x_{it} evolve stochastically, then the expectation has to be taken over the distribution of future values.

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

Estimation of the dynamic model requires that the researcher has data on the children's work experience, h_{it}. More generally, assume that the researcher has longitudinal data and denote t_{1i} and t_{Li} as the first and last periods of data observed for family i. In that case, the likelihood function is

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

where

$$\Pr(d_{i\tau}=1, w_{i\tau}) = \Pr(\xi_{i\tau} \geq -\xi_{i\tau}^{*}(\Omega_{i\tau}^{-}), \eta_{i\tau} = w_{i\tau} - z_{i\tau}\gamma_1 - \gamma_2 h_{it}) \text{ and}$$
$$\Pr(d_{i\tau}=0) = 1 - \Pr(\xi_{i\tau} \geq -\xi_{i\tau}^{*}(\Omega_{i\tau}^{-}))^{9}.$$

Identification requires the same exclusion restriction as in the static case, that is, the appearance of at least one variable in the wage equation that does not affect the utility value of school attendance, i.e., that does not appear in x. Work experience, h_{it}, would serve that role if it does not also enter into x_{it}. Note that the difference in the future component of the expected value functions under the two alternatives in Equation 19 is a nonlinear 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 Equation 19 as

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

where $G(\cdot)$ is the difference in the future component of the expected value functions, the nonlinearities in G that arise from the distributional and functional form assumptions may be sufficient to identify the discount factor.

Analyzing the impact of the policy experiment of introducing the school attendance subsidy is not any different in the dynamic than in the static setting. If the critical value of ξ that determines the choice of working or attending school is $\xi_{i\tau}^{*}(\Omega_{i\tau}^{-})$ without the subsidy, it is $\xi_{i\tau}^{*}(\Omega_{i\tau}^{-}) - \tau$ with the subsidy. The increase in school attendance is thus $\Phi(\frac{-\xi_{i\tau}^{*}(\Omega_{i\tau}^{-})+\tau}{\sigma_\xi}) - \Phi(\frac{-\xi_{i\tau}^{*}(\Omega_{i\tau}^{-})}{\sigma_\xi})$, again highlighting the importance of having the necessary exclusion restrictions to identify σ_ξ.

2.2.3. The multinomial choice problem. The binary choice problem considers two mutually exclusive alternatives, whereas the multinomial problem considers 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 Equation 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 $E[\max(V_t^0(\Omega_{it}), V_t^1(\Omega_{it}), \ldots\ldots, V_t^K(\Omega_{it}))]$ function that enters the multinomial version of Equation 18 at all values of Ω_{it}^{-} and at all t, adds significantly to that computational burden. If we define d_{it}^{n} as the discrete {0,1} choice

[9] If the structure does not yield an additive (composite) error, the latent variable function becomes $v_t^{*}(\Omega_{it}^{-}, \eta_{it}, \varepsilon_{it0})$. 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 $\text{Emax}(\Omega_{it}^{-})$ function must, in that case, be done numerically.

variable corresponding to the n-th choice (n = 1, . . ,N) and $\tilde{d}_{it}$ as the N element vector of those choices, there would be then at most $K = 2^N$ mutually exclusive choices.

Unobservable heterogeneity: permanent differences across agents that potentially affect the decisions that they make but that are unobserved by the researcher

2.2.4. 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, if a family was of type j, the preference for school attendance might be specified as $\alpha_{ijt} = \alpha_{oj} + x_{it}\beta + \varepsilon_{it}$ and the child's wage offer as $w_{ijt} = \gamma_{oj} + z_{it}\gamma_1 + \gamma_2 h_{it} + \eta_{it}$.[10] A type j family would be distinguished by the $(\alpha_{oj},\gamma_{oj})$ pair. Thus, potentially, families who value child schooling more might also have children who are more (or less) 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.

2.2.5. More flexible specifications. Estimable DCDP models are not nonparametrically identified (Rust 1994b) and are usually fully parametrically specified. However, any DCDP that can be numerically solved can, in principle, be estimated, and the DCDP structure does not restrict the choice of functional forms for preferences, technologies, and institutional constraints (e.g., tax rules), including the way in which unobservables enter (e.g., nonadditive errors, serial correlation), nor does it restrict the distributional assumptions of unobservables. The restrictions that are typically made arise from practical considerations, e.g., about the size of the state space and the number of parameters that must be estimated.

3. APPLICATIONS

In this section, we review five prominent and challenging social/economic policy issues in development economics in which the DCDP approach has been applied. These papers also illustrate the variety of structures that the DCDP approach can accommodate.

3.1. What Policies Are Effective in Increasing Educational Attainment and/or Improving School Quality in Developing Countries?

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 the papers reviewed in this section, DCDP models are used to evaluate the impact of incentives provided to families to increase their children's school attendance, the impact of incentives provided to teachers to reduce their absentee rates, and the impact of a school voucher program.

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

Conditional cash transfer programs: programs that give cash transfers to families or individuals conditional on their engaging in some behavior, such as attending school or visiting a health clinic

3.1.1. Mexico: Todd & Wolpin (2006) and Attanasio et al. (2005).

In 1997, the Mexican government introduced a conditional cash transfer program in rural areas that provided a subsidy to families for each child that met a school attendance goal. The initial program, called PROGRESA, has since been extended to urban areas in Mexico (and renamed Oportunidades), and similar programs have been adopted in numerous other countries (for example, in Bangladesh, Brazil, Colombia, Guatemala, and Pakistan).

To evaluate the initial 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 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 (Schultz 2004, Behrman et al. 2005).

An important limitation of large-scale social experiments, such as PROGRESA, 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. In the PROGRESA experiment, all eligible treatment group households faced the same subsidy schedule, so it is not possible to evaluate the effects of alternative subsidy schemes 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.

Todd & Wolpin (TW) (2006) and Attanasio et al. (AMS) (2005) analyze the impact of the PROGRESA program on school attendance via the estimation of a DCDP model of decision making about children's schooling. They use the model to compare the effects of the existing subsidy program with the effects of various alternative (nonexistent) programs. Although both adopt the DCDP approach, the models differ nontrivially in their structure. The two papers, however, use the same data set (although with different sample selection rules) and general empirical approach and perform similar counterfactual exercises. To the extent that their findings are quantitatively close, the conclusions would presumably be more credible.

We first provide a general description of the PROGRESA data and then describe the two models, their different approaches to using the data, and their empirical findings. 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 approximately 25% of family income. The school attendance subsidy component amounted to approximately 75% of total payments.

The school attendance subsidy schedule is shown in **Table 1**. 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 is larger for girls in grades 7–9 because of the greater tendency for girls to stop at grade 6.

Table 1 Monthly transfers for school attendance under the PROGRESA program[a]

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

[a]Corresponds to 1998–1999 school year. Table taken from Schultz (2004, table 1).

In the TW model, each year a married couple decides whether each of their children between the ages of 6 and 15 will attend school, remain at home, or, for children age 12 to 15, work in the labor market. These are assumed to be mutually exclusive activities. The couple also decides whether the wife will become pregnant (while fecund). The couple receives utility in each period from their current stock of children, their children's current years of schooling, their children's current school attendance, 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 time-varying shocks.

The household's income includes the parent's income and the wage income of the children who work, both of which are subject to time-varying shocks. A child's wage (offer) also depends on the child's age and sex, the distance to the nearest city, and household type. The parameters of the model are estimated by simulated maximum likelihood.

The AMS model also includes the binary choice of school or work (excluding the at-home option), but unlike the TW model, it assumes that each child's utility is maximized independently of that of the parents or of other children. The school/work decision is made at each age from 6 to 17, at which time there is a terminal payoff that depends on the number of years of schooling completed. The child receives a wage offer in each period that is village/education/age-specific. If the child rejects the wage offer and instead attends school, the child receives a utility payoff (positive or negative) that depends on observable preference shifters (parental background, the child's age, and the state of residence), the number of years of past attendance, observable variables that affect the cost of attending primary or secondary school (distance to a secondary school), on a child's unobserved discrete preference type, and on a time-varying preference shock that is assumed to be distributed as extreme value.[11]

[11]The extreme value error assumption implies that the dynamic programming problem as well as the choice probability (conditional on the wage offer) have closed-form representations (Rust 1987). To exploit this feature of the extreme value assumption, AMS estimate the wage function separately from the school/work decision rule.

In addition, AMS allow for a direct effect of the program on the utility of school attendance, either a feel-good effect from participating in the program or a stigma effect. The possibility that there may be an intrinsic value of participation in the program per se requires that both the treatment and control households are used in estimation. In contrast, TW hold out treated households from the estimation sample, using these households for external validation of the model.

TW compare the predicted effects on completed schooling of the PROGRESA program, as implemented, with that of alternative programs. As the first column in **Table 2** shows, the model predicts, based on simulations of households from the time of marriage until the last-born child reaches age 16, 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 grade 9. The first counterfactual (second column), the perfect enforcement of a compulsory school attendance law, establishes a maximal potential program impact. 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 third column shows the predicted effects of the PROGRESA subsidy schedule. The model predicts an increase in completed schooling of approximately one half-year for both boys and girls, or 25.9% of the maximal potential increase for girls and 28.0% for boys.[12] 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 26,096 (1997) pesos.

As noted above, the PROGRESA subsidy schedule rewards school attendance starting at grade 3. However, attendance in grades 3–5 is almost universal, making this aspect of the program essentially a pure income transfer. TW calculated that the per-family cost of the program could be held roughly constant if the subsidy in grades 3–5 were eliminated and the subsidy in grades 6–9 were increased by approximately 45%. The fourth column of **Table 2** shows that the cost of the grade 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 grade 9 increases by 3.4 percentage points and the proportion of boys by 3.8 percentage points, although there was a small decline in the proportion of children who complete at least grade 6.

Alternative programs, shown in **Table 2**, are (*a*) a 3000-peso graduation bonus for completing grade 9, which has a smaller impact on average completed schooling but a larger impact on the grade 9 completion rate and a per-family cost that is substantially more than the original subsidy, and (*b*) building a junior secondary school in 74% of the villages that do not have one, which has a similar effect as the modified plan (fourth column in **Table 2**) in terms of completed years of schooling.[13]

AMS perform two counterfactuals. As in TW, (*a*) they simulate the impact of eliminating the subsidy to primary school and redistributing the savings to increase the subsidies at later grades, and (*b*) they simulate the impact of building schools. Like TW, they find the effect of the first experiment to be large. However, the metric used by AMS is not directly comparable with that of TW. They find that the budget-neutral effect of eliminating the

[12]Interestingly, this estimate corresponds closely with that obtained by Behrman et al. (2005) and Schultz (2004) using nonstructural approaches.

[13]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.

Table 2 The effectiveness and cost of alternative education policies[a]

	Baseline	Compulsory attendance	PROGRESA subsidy	Modified subsidy	Bonus for completing grade 9	Build schools
Mean completed schooling						
Girls	6.29	8.37	6.83	6.97	6.50	6.39
Boys	6.42	8.29	6.96	7.07	6.58	6.55
Percent completed grade 9						
Girls	19.8	55.5	25.9	29.3	28.8	21.2
Boys	22.8	54.7	28.0	31.8	32.7	24.1
Cost per family[b]	0	Unknown	26,096	25,193	36,976	Unknown

[a]Table taken from Todd & Wolpin (2006).
[b]Cost given in 1997 pesos.

subsidy at younger ages increases school attendance rates by as much as 100% at age 15–16. Also consistent with TW, AMS find a large effect of building schools on the school attendance of older children as in their modified program. These findings are, perhaps, surprisingly robust given the quite significant differences between the model structures and estimation samples of TW and AMS.[14]

In assessing the value of the DCDP structural approach, it is useful to recall the finding in both papers about the effect on schooling of a budget-neutral shift in resources toward the higher grades. That such a shift would increase schooling overall must be true from the fact that the attendance up to grade 5 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 all of them.

3.1.2. India: Duflo et al. (2008). Although increasing school attendance among children is a necessary prerequisite for increasing school attainment, having a teacher in the classroom would also seem to be an equally necessary input. Duflo et al. (DHR) (2008) study a program that provided monetary incentives for teachers in rural India to reduce teacher absenteeism. Like the PROGRESA program, the incentive system was introduced as a randomized field experiment. The program was implemented by a nongovernmental organization and began in August 2003. It randomly selected 120 schools, with half assigned to the control group and half assigned to the treatment group, in which teacher attendance was monitored on a daily basis. Each teacher was given a camera and was asked to have a

[14]AMS directly incorporate general equilibrium effects of increased school attendance on wages. Although they find that child wages rise by 6% due to the program, they also find the attenuation effect on school attendance to be small. Their result is consistent with the counterfactual performed in TW, who find the attenuation effect to be moderate if wages were to increase by 25%.

picture taken at the beginning and end of each school day. To demonstrate that the teacher was not absent, the time- and date-stamped picture taken by one of the students had to include the teacher and at least eight students. In addition to the camera-based evidence, there was a random spot check of teacher attendance in each school once per month.

When the program began, the monthly teacher salary, based on at least 20 days of work, was Rs. 1000. The treatment group was paid a Rs. 50 bonus for each day over 20 that the teacher was present and fined Rs. 50 for each day below 20 (with a maximum fine of Rs. 500); the monthly pay of teachers in the treatment schools ranged from Rs. 500 to Rs. 1300. The control group was paid a flat Rs. 1000 for 20 days of work and was told that they could be fired for excessive absences. The daily attendance rate at the start of the program was approximately 65%. Between September 2003 and February 2006, the teacher attendance rate averaged 79% in the treatment schools, but only 58% in the control schools, indicating a pronounced effect of the program on teacher absenteeism.

To further assess the impact of the incentive program and of variations to the incentive schedule, DHR specify and structurally estimate a DCDP model of teachers' daily decision of whether to work or be absent. The utility of working on a given day in a treatment school is the increment to the wage that the teacher will earn by the end of the month, which depends on how many days the teacher has worked so far in the month. A teacher receives Rs. 500 for the first day of work, a zero increment up to 10 days of work, and an extra Rs. 50 for each day of work above 10. A teacher who does not work receives a stochastic utility flow with a given mean. DHR assume that a teacher in a control school receives Rs. 1000 regardless of the days worked, so her daily choice is governed only by whether her utility from not working is positive. There are assumed to be two types of teachers that differ by the value placed on not working.

The model is estimated using the treatment schools only. As in TW, the control schools are used for out-of-sample validation and for model selection, where the modeling choice is between i.i.d. and serially correlated error structures. DHR find that the model with serial correlation better fits the data in both the treatment and the held-out control schools.

DHR perform counterfactual experiments that vary the payment schedule, that is, the minimum number of days worked before being eligible for the bonus and the size of the daily bonus, to determine the lowest cost schedule that would result in a given minimum expected number of days worked. **Table 3** reports the estimates from those counterfactual exercises. As the table shows, with no bonus, that is, with a flat payment of Rs. 500, the expected number of days worked would be 14. Providing a bonus of Rs. 25 for each day after having worked 21 days would achieve an expected number of work days of 15 at the least cost. The incremental cost of the program for an additional day of work per teacher would be Rs. 21. The cost for further additional days increases at a nonlinear rate, reaching Rs. 2624 to achieve 21 days and Rs. 4604 to achieve 22 days. At the current program cost, using the efficient payment schedule would have increased the number of the days worked to 20, a 16% increase for the treatment group.

3.1.3. Chile: Bravo et al. (2009). School vouchers were first proposed by Friedman (1955, 1962) as a way of improving the quality of schooling and increasing opportunities for disadvantaged children. The question of whether school vouchers would improve educational quality has been the subject of intense debate in the United States. Most of the school voucher programs that have been implemented in the United States are small-scale experiments, so the U.S. evidence on the effectiveness of a universal voucher program is sparse.

Table 3 Counterfactual cost minimizing policies[a]

Expected days worked	Bonus cutoff	Bonus	Expected cost[b]
14	0	0	500
15	21	25	521
16	22	75	664
17	21	75	672
18	20	75	755
19	20	100	921
20	20	125	1112
21	16	225	2642
22	11	275	4604

[a]Table taken from Duflo et al. (2008).
[b]Cost given in 2003 rupees.

Chile is one of the only countries in the world that has long-term experience with a nationwide school voucher program, having adopted a universal program in 1981 that is still operating today. At its inception, the control of public schools was transferred to municipal authorities, and the school funding system was converted to a per-capita voucher system, with public and private schools receiving the same voucher amounts.[15] In the first five years after the voucher program was introduced, the percentage of students enrolled in private subsidized schools increased rapidly, from 15% to over 30%, with a corresponding decline in public school enrollment. Today, more than half of primary and secondary school students in Chile attend private schools.

School vouchers: a designated amount of funding that a student can use to pay for school tuition at a school of choice

Education in Chile is provided by three broad types of schools: municipal schools, private subsidized schools, and private nonsubsidized (fee-paying) schools.[16] Private nonsubsidized schools, which include both religious (mainly Catholic) and lay schools, are financed from private tuition and are often for profit. Parents are free to choose among municipal and both types of private schools. Private schools can be selective, however, whereas public schools can be selective only if there is excess demand. In all types of schools, students take standardized tests, called the SIMCE tests, and summary measures of test results by school are publicized and used by many parents as an indicator of school quality.

To study how the educational reforms in Chile affected school choice, longer-term educational attainment, and labor market outcomes, Bravo et al. (BMT) (2009) develop and estimate a dynamic behavioral model of schooling and labor force participation decisions. The analysis samples come from the 2002 and 2004 waves of a longitudinal survey in Chile, the Enquesta Proteccion Social (EPS), which elicited information from

[15]This change greatly increased the level of public funding going to private schools, roughly doubling it, but it did not alter overall per-pupil spending, which actually declined somewhat in the decade of the 1980s.

[16]Until 1994, private subsidized schools and municipal schools were financed primarily through the per-capita government voucher. In 1994, there was a change in rules to allow public and private schools to impose a small tuition charge on top of the voucher. Municipal schools sometimes also receive some additional funding in the form of government transfers when the voucher amounts are not sufficient to cover the school's operating expenses.

respondents on the primary and secondary schools attended and on educational and labor market outcomes, including a retrospective work history going back to 1980.[17] A major challenge in estimating the effects of school vouchers is that the voucher program was introduced throughout Chile in 1981 with no variation in the timing of availability. However, there is variation in exposure to the program that can be exploited, given that Chileans were at different points in their schooling career when the reforms were introduced.

In their DCDP model, decision making starts at age 6 and ends at age 65. At ages 6–15, individuals decide sequentially whether to attend school or stay at home. At ages 16 and older, individuals receive wage offers each period and decide whether to attend school, stay home, or work. The model assumes that children must remain in school through grade 4 and that it is not possible to return to school after leaving. An individual who graduates from high school (grade 12) has the option of attending college for up to five years. With regard to the type of schooling choice, individuals choose whether to attend public, private subsidized, or private nonsubsidized school, which is a one-time choice when starting primary school (grade 1) and when starting secondary school (grade 9).

Wage offers, reflecting an individual's stock of human capital, depend on the type of school attended, the number of years attended, whether the schools were attended during the pre- or postvoucher regime, labor market experience, and an individual's unobserved type.[18] There is also a net nonpecuniary value of attending school that is subject to a preference shock each period and that differs by the individuals unobserved type, by the type and level of school, and by whether the individual resides in Chile's capital city (Santiago). The value of attending different types of schools is allowed to vary pre- and postvoucher reform to capture changes in the cost of attending school due to the much wider availability of private schools after the reform. The utility of leisure depends only on an unobserved type-specific intercept and a leisure preference shock.

The model, which is estimated by simulated maximum likelihood, is used to assess how the school voucher reform influenced sorting among different types of schools, educational attainment, earnings, and labor market participation. By simulating choices over the life cycle with and without the voucher reform, it is possible to evaluate the cumulative effects of the voucher reform as it operates through schooling and labor market channels. Specifically, the voucher reform modifies the budget constraint by eliminating tuition at private subsidized schools, the net costs of attending different types of schools, and the returns to schooling.

The parameter estimates indicate that the return to all types of primary schooling increased after the voucher reform, with primary municipal schooling exhibiting the largest increase. Such an increase is consistent with increased competition having improved the quality of all types of primary schools. Competition tends to be strongest at the primary school level, where there are a large number of schools. At the secondary school level, however, the postvoucher estimated wage returns to secondary schooling are somewhat lower than prevoucher levels, which likely reflects that per-pupil expenditure declined over the decade following the voucher reform and that school competition is potentially less significant at the secondary level, where there are fewer schools. For both primary and

[17]The first round of data was collected under the survey name Historia Laboral y Seguridad Social.

[18]The model evolves with age, not calendar time. However, to allow for nonstationarity in wages over time, type proportions are allowed to depend on 10-year birth cohorts as well as family background characteristics.

secondary schools, however, the estimated costs of attending school declined significantly after the reform.

On net, BMT find that the voucher reforms increased primary school graduation rates by 0.6 percentage points, high school graduation rates by 3.6 percentage points, college-going rates by 3.1 percentage points, and the percent completing at least four years of college by 1.8 percentage points for individuals exposed to the reform during their entire schooling career. In addition, the reform reduced labor force participation at ages 16–25 by approximately 2 percentage points, off a baseline of 58.3 percentage points, because longer school-going delays labor force entry. Perhaps surprisingly, the voucher reform did not lead to increased overall mean wages, in part because the wage benefits of having more education are partly offset by the postreform decrease in the returns to secondary schooling. However, the variance in earnings decreases, because program impacts on earnings are relatively large and positive at lower percentiles of the distribution and negative at the higher percentiles. The impacts of the voucher reform are similar in magnitude for individuals from both poor and nonpoor backgrounds.

3.2. How Do Government Pension Programs Affect Household Labor Supply and Retirement Decisions?

Many developing countries have few formal mechanisms for providing old-age pension support or have formal pension programs that only cover a small proportion of the population. A few DCDP models, reviewed below, have been developed for the purpose of studying how the design of a pension program influences labor force sector participation decisions, pension program coverage rates, retirement savings, and transfers from other household members.

3.2.1. Indonesia: McKee (2006). Most developing countries have few formal mechanisms in place for providing old-age pension support. In the absence of formal pension systems, older individuals often rely on their own labor earnings, coresidence with other household members, or partnerships in family businesses to support their consumption. Indonesia, the fourth largest country in the world, has been experiencing rapid demographic changes that affect the sustainability of these kinds of traditional forms of old-age support. Improved provision of water, sanitation, and health care led to a substantial increase in life expectancy, and the widespread availability of family planning programs contributed to a decrease in fertility. Both factors contribute to increased dependency ratios.[19] The Indonesian government is therefore interested in understanding what effects changing demographics will have on the well-being of the elderly and in developing other means of providing support for old-age individuals, such as expanding the coverage of existing old-age pension programs that are currently available only to a small fraction of workers.

To explore these issues, McKee (2006) develops and structurally estimates a DCDP model of household labor supply for males age 40–75 tailored to the Indonesian institutional setting. Currently, civilian and military workers in the public sector have access to a public pension program that requires a contribution of approximately 4.75% of their salary. After age 55 for civilians (50 for the military), they qualify for a public pension

[19]Completed fertility declined from 6 to 2.2 children on average. In Indonesia, more than half of couples with adult children currently receive monetary transfers from at least one child.

Privatized pension funds: pension funds that an individual keeps in a private pension account, usually of his/her choosing

with (inflation-adjusted) benefit levels that depend on their last month's salary and length of service. Private sector workers have access to a less generous defined contribution program. Participating workers contribute 2% of their salary, and employers contribute an additional 3.7%. At age 55 retiring workers collect the amount contributed plus the investment return, which has historically been approximately 0% in real terms. For this reason, there is very low take-up in the private pension program.

McKee's model incorporates uncertainty about future health, wages, and other income. It also accounts for family structure and allows for health- and age-related declines in wages. In the model, the male household head maximizes his expected discounted lifetime utility. At each age between 40 and 74, an individual probabilistically receives job offers consisting of particular (weekly) hours-sector combinations. Hours offers are restricted to be either 20, 40, or 60 h, and job offers come from either the private or government sectors or as a self-employment opportunity. If an individual receives no offer, the choice set is 0 h. Limited availability of jobs helps to explain why individuals do not always choose a government sector job, even though those jobs generally pay better and provide a pension. Individuals must retire by age 75 and live to at most age 85. At ages less than 85, the individual faces a known probability of survival.

Own monthly labor income depends on weekly hours worked and on an hourly sector-specific wage, which also depends on health status, education, age, and weekly hours of work. The male household head acts as a Stackelberg leader with respect to the labor supply decisions of the other household members and with respect to monetary transfers from family members living outside the household. The head receives given fractions of income generated by the labor supply of other members and from transfers. The model also incorporates the features of Indonesia's government employee (public sector) pension system but not the private sector pension program, because of its very low take-up.

The data used to estimate the model come from the 1993 and 1997 waves of the Indonesian Family Life Survey, which covers 13 of Indonesia's 26 provinces and is representative of 83% of Indonesia's population. The analysis sample consists of a sample of 2993 men between the ages of 40–75 for whom information on demographics, work experience, health, assets, and transfer payments was reported. The data set contains retrospective information on wages, hours, and job sectors going back to 1988. The health measures in the survey include general measures of self-reported health and standard measures of activities of daily living.

The model is estimated by simulated maximum likelihood. McKee considers three different counterfactual policies: (*a*) a substantial reduction in family support for the elderly, (*b*) an unfunded expansion of the government pension program into the private sector, and (*c*) the same pension expansion under the assumption that the government imposes a tax on wages of private sector workers to pay for the program. With regard to the first counterfactual, McKee finds that halving the amount of support that individuals receive from their family increases the labor force participation of older individuals by approximately 5%. Most of the increase in employment would take the form of self-employment. Under the unfunded pension reform, the availability of a private pension induces individuals to increase their labor supply and switch to working in the private sector. Under the fully funded pension reform, before age 55 labor force participation increases in a similar way as observed for the unfunded pension reform simulation, but after 55, participation in the private sector declines significantly when individuals must pay the wage tax. From these simulations, McKee concludes that family demographics and health play key roles in labor supply choices and that older men would work longer if family support declines. Also, the availability of better pension benefits in private sector jobs would significantly affect the choices that individuals' make between unemployment, self-employment, and private sector work.

Counterfactual policies: policies that have not been implemented before and are therefore not within the realm of historical experience

3.2.2. Chile: Vélez-Grajales (2009). Many pay-as-you-go social security systems in the United States and Europe face impending insolvency as the number of pensioners per worker rises. One type of reform being considered is transitioning to a fully funded pension system. Chile has been at the forefront of pension reform, having switched to a fully funded private retirement accounts system 28 years ago. The Chilean pension system has served as a model for pension reform in many Latin American countries and is also being considered as a possible prototype for reform in the United States and in Europe. However, it has also come under criticism, for example, for having low coverage rates. The Chilean government recently implemented changes to the system aimed at increasing coverage and expanding the social protection components. Vélez-Grajales (2009) develops and estimates a DCDP model to evaluate how the design of Chile's pension system influences the system's performance, as reflected in worker's labor force participation and pension contribution decisions.

As in many Latin American countries, Chile's labor market has both formal and informal sectors. In the formal sector, workers pay taxes and are mandated to participate in the public pension and health plans. The pension plan requires that they contribute 10% of their monthly earnings to a pension management fund of their choice, along with an additional percentage for fees and commissions. The legal age of retirement in Chile is 65 for men and 60 for women, but workers can retire earlier if they have sufficient pension accumulations. Workers who have contributed for 20 years but do not have sufficient accumulations qualify for a minimum pension benefit guarantee. In the informal sector, much of the income is undocumented, but workers are able to contribute voluntarily to a pension plan. There is also no mandate for self-employed workers to contribute, but one of the reforms being considered is to make contributions mandatory for self-employed workers. Workers who have not contributed to a pension plan are eligible for a welfare pension upon retirement.

To study how the design of the pension system influences working and pension contribution decisions, Vélez-Grajales develops a DCDP model of individual decision making about labor force participation and pension contributions. Her analysis focuses on men, because men have much higher rates of labor force participation than women in Chile. Vélez-Grajales' model begins at the age when an individual finishes his studies (age 18 for those who do not go to college) and ends at age 85. An individual makes decisions until retirement. In each period, which corresponds to one year, a worker receives wage offers from both the covered and uncovered sectors and decides where to work and how many quarters within the year to work. An individual may also decide to retire, that is, to stop working permanently and collect the pension. In the uncovered sector, workers also decide how many quarters to contribute to the pension plan from the number of quarters that they work. Health status and marital status evolve exogenously.

The empirical analysis is based on panel data from the 2002 and 2004 waves of the EPS, the same data used in the BMT school voucher study. The survey includes retrospective data on employment, nonemployment, and unemployment back to 1980, as well as wage data for 2002 and 2004. The EPS data are linked to administrative social security records back to the date at which the individual entered the privately funded pension system (or

1980, which is the earliest date) that include mandatory and voluntary pension contributions, monthly wages, and choices regarding pension fund administrators and pension plan investments. The analysis sample is composed of 2084 men age 18–39 years old who entered the labor force after 1981 and are known to have contributed only under the new pension system and not under the old pay-as-you-go pension system. The parameters of the model are estimated by simulated maximum likelihood.

The estimated model is used to perform three policy experiments. As noted above, a major concern of policy makers with regard to the Chilean pension system is how to increase coverage rates. One policy experiment reduces the number of years required for the minimum pension from 20 to 15 or 10 years. A second policy experiment increases the level of the minimum pension by 10% and 25%, as an alternative way of providing inducement to participate in the pension program and as a way of expanding the social insurance feature of the pension program. A third policy experiment explores whether the level of commissions charged to manage pension funds discourages participation in the pension program, as some policy makers have argued, by decreasing the average level of commissions from 1.7% to 1.2% and 0.7%. Because the behavioral model was estimated on a sample of relatively young men, the policy experiments focus primarily on the effects on labor supply and contribution rates at younger ages, but also consider longer term effects on retirement behavior, which requires an extrapolation outside the range of the data.

Table 4 shows the average number of accumulated quarters spent working in the covered and uncovered sectors and not working under alternative required numbers of quarters of participation needed to qualify for the minimum pension. For males age 39, reducing the requirement from 80 to 60 quarters would increase quarters of work in the covered sector by 0.9 and reduce quarters of work in the uncovered sector and in leisure by 0.4. The effects increase in magnitude when the requirement is further reduced to 40 quarters. The estimates indicate that relaxing the requirement on the number of quarters required to obtain a minimum pension would increase covered sector work, although the estimated response is fairly modest. The model simulations also show that decreasing the number of quarters required to get a minimum pension benefit is costly for the government because it induces people to retire earlier and to opt for the minimum pension benefit.

Table 4 Accumulated mean quarters of work by labor market sector at baseline (BL) and under counterfactual experiments varying required quarters of contributions[a]

	Labor market sector								
	Covered			Uncovered			Out of work		
Age group	80 BL	60 Exp. 1	40 Exp. 2	80 BL	60 Exp. 1	40 Exp. 2	80 BL	60 Exp. 1	40 Exp. 2
24	5.6	5.9	6.6	10.8	10.8	10.6	6.8	6.4	5.9
29	14.0	14.5	15.6	19.8	19.6	19.2	7.7	7.3	6.7
34	23.3	24.0	25.3	29.6	29.3	28.6	8.6	8.1	7.5
39	32.6	33.5	34.9	39.3	38.9	38.1	9.5	9.1	8.4

[a]Table taken from Vélez-Grajales (2009, table 7.1).

The policy experiment in which the size of the minimum pension is varied indicates that an increase in the size of the minimum pension by 25% would lead to a substantial increase in accumulated quarters of pension contributions. As seen in **Table 5**, by age 39, accumulated quarters of contributions would be three times greater than under the current program. In addition, people also choose to retire earlier, and a greater proportion retire with a minimum pension. This happens because, with the higher rate of contributing, individuals tend to fulfill their minimum pension requirement much earlier, in their fifties instead of in their sixties. Although the percentage of individuals who retire with a minimum pension is similar under this policy to that when the required contribution is reduced to 40 quarters, the cost to the government of providing the minimum pension benefit, that is, of making up the difference between the actual benefit the individual has and the minimum guarantee level, is much higher under the policy that increases the minimum pension benefit size. Lastly, the third policy experiment that decreases the level of commissions charged by the pension plan management agencies (not shown in the tables) indicates that commissions and fees in the range that are currently charged are not a major factor determining pension program participation.

Microfinance programs: programs that provide cash or in-kind transfers (as grants or loans) to entrepreneurs or family enterprises

3.3. What Policies Are Effective in Increasing Business Investments by Households in Developing Countries?

Credit constraints are thought to be a potentially major factor inhibiting economic growth in developing countries. In this section, we discuss two papers that provide evidence about the importance of such constraints in limiting productive capacity. The first paper considers the extent to which productive assets are used for the purpose of consumption smoothing in an agricultural setting and quantitatively assesses policies that may improve productive efficiency. The second paper analyzes the impact of a large-scale microfinance program on stimulating investment behavior by households and assesses its efficacy relative to a direct transfer program.

3.3.1. India: Rosenzweig & Wolpin (1993). Many of the assets held by Indian farmers are also productive inputs. Given the existence of markets for those assets, they can also be used as a store of wealth or buffer stock to smooth consumption in the face of adverse income shocks when credit markets are imperfect. Rosenzweig & Wolpin (RW) (1993)

Table 5 Accumulated quarters of contribution in the uncovered sector and in both the covered and uncovered sectors when the size of the minimum pension is varied relative to baseline (BL)[a]

	Uncovered			Both sectors		
Age group	0.8 BL	0.88 Exp. 1	1.0 Exp. 2	0.8 BL	0.88 Exp. 1	1.0 Exp. 2
24	1.8	2.3	5.3	7.4	7.9	12.9
29	2.8	3.6	9.5	16.8	17.6	26.4
34	3.2	4.2	12.4	26.4	27.5	39.2
39	3.3	4.3	13.7	35.9	37.1	50.3

[a]Table taken from Vélez-Grajales (2009, table 7.6).

document that bullock holdings are a major source of wealth among Indian farmers and that they respond to both positive and negative income shocks. Distress sales, in particular, would not arise if consumption could be smoothed either through the sale of nonproductive assets, such as jewelry, or through borrowing. Bullock markets are, on the other hand, well developed and spatially integrated, enabling farmers to move bullocks between areas that face different weather shocks. Indeed, RW report that bullock sales were 3.5 times more prevalent in their data than was the sale of jewelry.

RW develop and implement a model that captures many important features of the low-income agricultural environment faced by Indian farmers. The model assumes that (*a*) farmers cannot borrow to smooth consumption; (*b*) there is no rental market for bullocks, although bullocks can bought and sold, and also bred; (*c*) water pumps can be bought and sold, but not rented; and (*d*) land cannot be bought or sold, but can be rented.

In each year, a farmer chooses whether to own zero, one, or two bullocks, whether to breed a bullock (which becomes productive three years later), and whether to make an irreversible investment in a pump. Profits depend on the number of bullocks, whether there is a pump, the farmer's age, an i.i.d. village-level weather shock, and an idiosyncratic time-varying i.i.d. shock. Flow utility is CRRA in consumption above some minimum level, and consumption is profits minus the cost of bullock purchases (or plus the revenues from sales), minus the cost of purchasing a pump and the breeding cost. An adult bullock may die in any period with a given probability.

The model is estimated on a subsample of midsize farmers taken from a longitudinal survey over the years 1975–1984 of 30 farmers in each of 10 villages located in five districts in the semiarid tropics of India. The two other groups of farmers, those with small and those with large land holdings, were used for out-of-sample validation. There are 33 farmers, chosen from three of the villages, and 263 farmer-year observations in the estimation sample. Within that group, bullocks are either bought or sold in 22.5% of the sample years, and almost 90% of the farmers bought or sold a bullock in at least one year.

Estimates of the profit function reveal that it is optimal (that is, profits are higher) to have a pump, which increases mean profits by 72%. However, only 31% of farmers ever owned a pump, and only 18% purchased a pump over the 10-year period. The profit function estimates also reveal that it is optimal to own two bullocks, whereas, on average, farmers own less than one bullock. Owning two bullocks, relative to none, leads to a similar increase in mean profits as owning a pump. The rest of the parameter estimates reveal that (*a*) flow utility is approximately logarithmic in consumption net of the consumption floor, (*b*) the consumption floor is approximately one-half of the level of mean household food consumption, and (*c*) the cost of breeding a bullock is approximately 86% of the price of an adult bullock.

As noted above, there is significant underinvestment in both pumps and bullocks, presumably because of the lack of credit availability and the inability of farmers to accumulate financial assets given their income flows. RW assess a number of potential policies that might ameliorate this underinvestment. The first policy they assess is the provision of actuarially fair weather insurance. They note that in a simulation in which farmers are initially provided with two bullocks (the optimal number), over time the occurrence of bad weather shocks, which cause distress sales of bullocks, is frequent enough that, given the cost of breeding and the bullock price, it is not optimal to hoard bullocks (that is, to hold two or more bullocks). It is thus possible that weather insurance, by smoothing income, will be welfare improving. RW find, however, that actuarially fair weather insurance, while smoothing consumption, also leads to lower average consumption given the cost of the insurance and leads to no welfare gain. They argue that the existence of the consumption floor, which reflects alternative informal risk-sharing arrangements (e.g., through family transfers), is close to a perfect substitute for weather insurance.

Because it appears that it is the combination of the lack of credit availability combined with low income that leads to inefficient bullock holdings, RW also perform an experiment in which farmers are guaranteed additional income. They find that a 50% supplement to farmers' annual income in each year would increase bullock stocks at age 50 by approximately 75% and to within 22% of the profit maximizing level of two bullocks. Also, at the higher incomes, farmers are more willing to accept income fluctuations, that is, to be less prone to distress sales of bullocks. RW conclude that providing additional earnings opportunities to low-income farm households would have the effect of improving agricultural efficiency.

3.3.2. Thailand: Kaboski & Townsend (2007). Microfinance programs are viewed as an important mechanism for stimulating investment in developing countries. However, there are few estimates of the economic returns from such programs or of how the returns compare with those from other types of programs, such as direct transfer schemes. To provide evidence on the relative effectiveness of microfinance programs, Kaboski & Townsend (2007) (KT) develop and structurally estimate a model of credit-constrained households deciding on consumption, indivisible investment, and savings. As in the papers by TW and DHR, they hold out part of the data for external validation. In particular, they estimate the model using data collected prior to the introduction of a large-scale government microfinance program, the Thai Million Baht Village Fund program, and then validate the model using postprogram data.

The Thai Million Baht Village Fund program, begun in 2001, involved the transfer of one million baht (approximately $25,000) to each of almost 80,000 villages in Thailand to start village banks that would lend to village households. The program was rapidly implemented, making it unlikely that households could have anticipated it. The total amount of funding to each village was the same regardless of the size of the village, so village size provides a plausibly exogenous source of variation in the amount of credit increase per household. KT view the program as an exogenous quasi-experimental increase in credit. The analysis samples come from the Townsend Thai project, which gathered panel data on rural and semiurban households and businesses from 64 villages in four Thai provinces from 1997 to the present.

The model is based on the standard buffer stock model of savings behavior under income uncertainty, along the lines of Aiyagari (1994) and Deaton (1991), but with the additional feature that households face an investment option. Specifically, in the model, households start the first period with some level of permanent income and liquid wealth, and a potential investment project. Income is the product of permanent income, which follows a random walk with drift, multiplied by a transitory income shock. Each period, the household makes a decision about whether to undertake an investment project of a given size, which earns a rate of return higher than the rate of return on liquid savings. Liquid wealth is bounded below by a borrowing limit, which is a multiple of permanent income.

The household maximizes the expected discounted value of utility over an infinite horizon by making decisions on consumption, savings, and on whether to undertake the investment opportunity. The model allows for the possibility of default, if the agent experiences a sufficiently bad shock to permanent income. In that case, the agent is assumed to consume a minimum consumption level that is proportional to permanent income.

The setup of the model differs somewhat from the DCDP framework discussed above. Specifically, whereas the choice of undertaking the investment option is discrete, the savings decision is continuous, as is the state variable given by the level of liquid assets at the start of any period. One approach to the solution and estimation would be to discretize the choice of savings levels. Alternatively, KT solve the first-order condition for savings for a subset of state space values and use an interpolation procedure to obtain the optimal level of savings at other state points.

The behavioral model is estimated by the generalized method of moments using the first five years of pre-experiment data. The crucial data elements used in the analysis are information from households and local businesses on their consumption, income, investment, credit, liquid assets, and interest income from these assets, in addition to village population data. The measure of household consumption is calculated using detailed monthly expenditure data.

The validity of the estimated model is assessed by comparing the model's predictions of the effects of the Thai Million Baht Village Fund program on consumption, investment, and the probability of investing to the actual effects observed under the experiment. The program is introduced into the model as a reduction in borrowing constraints in an amount that would increase the amount of total expected credit (as calculated from the model) in the village by one million baht. Impact estimates obtained using the model's simulated data are very close and in fact not statistically different from impact estimates obtained from reduced form regressions based on actual postexperiment data. One of the notable predictions of the model that is also borne out in the data is that the impact on consumption exceeds one million baht.

The estimated model is used to compare the costs of the microfinance program with the costs of a direct transfer program that would provide the same utility benefit. They find that the cost of the microfinance program is 33% less, attributable to the fact that the microfinance program relaxes borrowing constraints, which the transfer program does not do.

Even households that do not use credit can be affected by the relaxation in borrowing constraints, as it lowers their need for a buffer stock of liquidity and allows them to invest and increase consumption. Households who increase their borrowing are those who have the highest marginal valuation of liquidity, which makes the village fund program more cost-effective than a simple transfer program.

The results indicate also that the largest effect of the microfinance program is on consumption rather than investment. This occurs because the program did not stipulate the use of the credit. KT thus perform a counterfactual that limits the use of credit to investment. Of course, the fungibility of the money implies that not all investments that use the program's credit are new investment, that is, investment that would not have been undertaken otherwise. Nevertheless, this restricted policy is found to increase the probability of investment by over three times relative to the unrestricted (actual) program and to be slightly more cost-effective. In summary, KT demonstrate that microfinance programs are an effective means of increasing liquidity of credit-constrained households, that they would positively impact both investment and consumption, and that they are more effective than a simple transfer program.

3.4. How Does Immigration Policy Affect the Flow of Migrants Between Developing and Developed Countries? Mexico/United States: Colussi (2006)

Network effects on migration: the influence of an individual's peer network on his/her decision to migrate to a new labor market

Mexican illegal immigration has been, and continues to be, a central U.S. policy concern. From 1942 to 1964, under the Bracero program, the United States provided a fixed annual number of temporary visas to Mexican agricultural workers. The number of visas peaked at approximately 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 the abandonment of 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 United States since 1982. A series of legislative and administrative initiatives was introduced throughout the 1990s aimed at reducing the flow of illegal immigrants, and new Congressional legislation has more 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 15–64-year-old potential migrants. A potential migrant chooses between working in the United States, 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 United States may entail a nonpecuniary cost (or benefit). The village economy is modeled as having a closed labor market, so that migrants to the United States reduce the supply of labor in their home village and put upward pressure on village wages.[20] Return migration occurs because the utility of residing in the United States declines with current duration in the United States, as well as from a response to transitory shocks to U.S wages.

Migrants from any single village have no effect on U.S. wages and take the (stochastic) wage process in the United States as given. However, potential migrants do not necessarily receive job offers. The probability that a potential migrant receives a wage offer in the United States depends both on his own personal characteristics, such as his U.S. work experience, and on his village network. The more migrants from his home village that are currently working in the United States, the higher the likelihood of receiving a job offer. The importance 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 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

[20]Colussi's is one of the few DCDP models in which policy evaluation is conducted within an equilibrium framework.

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 **Figure 1**. The initiation and later abandonment of the Bracero program are clearly seen, as is the subsequent rapid rise in migration, particularly in village 1, 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 2, 73% for village 3, and 85% for village 1.

Given the estimates of the model, Colussi simulates the effect of increasing enforcement on the flow of illegal immigrants. **Table 6** 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 U.S. wages to one month's wages and to six month's wages. In the second, migrant wages in the United States 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 U.S. wages would actually increase the fraction of migrants in the steady state by 9%. This increase arises because, although it is optimal for migrants to make fewer trips at this higher cost per crossing, the trips are of longer duration. The average duration of stay increases from five to seven 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 bring about a reduction in the steady-state fraction of migrants by approximately 12%. Thus, the steady-state fraction of migrants appears to be relatively invariant to the cost of crossing the border.

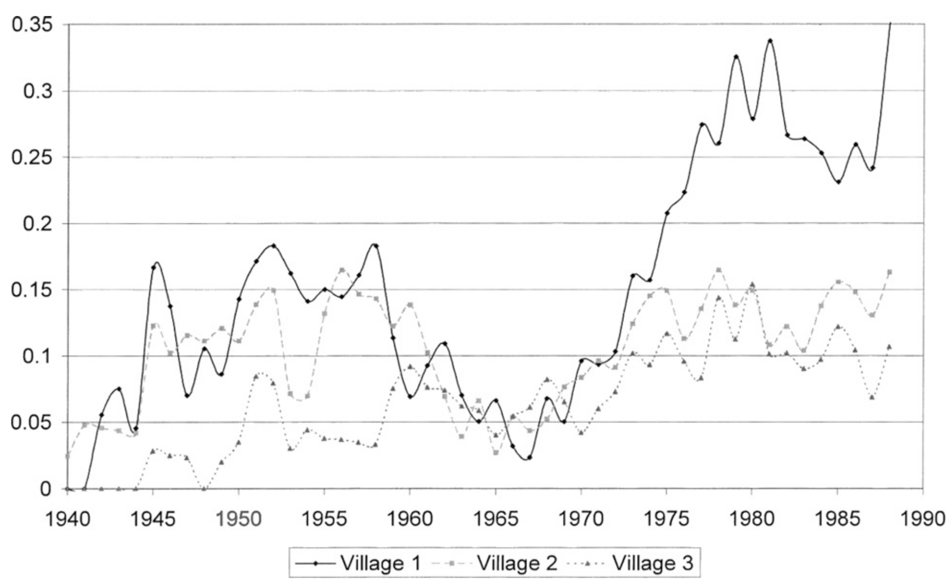

Figure 1

Percentage of Mexican villagers in the United States by year.

Table 6 The effect of increased enforcement on illegal immigration (village 1)[a]

		Cost of border crossing[b]		U.S. wage reduction	
	Baseline	One month U.S. wages	Six months U.S. wages	10%	25%
Steady state fraction in United States	0.32	0.35	0.28	0.27	0.02
Average duration in United States (years)	5	7	19	4	2

[a]Table taken from Colussi (2006).

Reductions in U.S. wages decreases both the fraction of migrants and their average length of stay in the steady state. A 10% reduction in the wage decreases the fraction of migrants by 16% and the average trip duration by 20%. Although 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. A direct comparison of the efficacy of these two policies is not possible without knowing the cost of their implementation, but 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. CONCLUSION

The applications summarized in this review were chosen to illustrate the breadth of challenging and important policy issues that have been addressed in the development economics literature using the DCDP approach.[21] The DCDP approach is not the predominant one in the development literature, although it is growing in popularity. In the past decade, the predominant methodology for policy evaluation in development was randomized field experiments. As illustrated by three of the papers summarized above (Attanasio et al. 2005, Todd & Wolpin 2006, Duflo et al. 2008), an emerging view is that there is a natural synergy between field experiments and structural approaches, like DCDP models.

By melding these approaches, the advantage of each can be exploited. If done well, a field experiment can identify as cleanly as possible the impact of a given policy on outcomes of interest.[22] However, field experiments are expensive, and it is not usually feasible to vary treatments very much within the experiment.[23] For example, the Mexican PROGRESA program studied by Todd & Wolpin and by Attanasio et al. and the Indian teacher incentive program studied by Duflo et al. had only a single treatment. The structural approach, however, can ameliorate this limitation by enabling counterfactual policy evaluation that goes beyond what is contained in a particular experiment.

The one additional requirement for combining these approaches is that the experiment include data beyond simple measurement of the treatment and the outcomes. The structural

[21]In a recent paper, Keane & Wolpin (2009) provide additional examples across a number of fields, including labor economics, health economics, and public economics.

[22]There is the caveat that field experiments provide evidence only on partial equilibrium effects.

[23]See Deaton (2009) for a discussion of the promise and limitations of field experiments in development economics. Also, see Heckman & Urzua (2009) for a discussion of the advantages and limitations of experiments and instrumental variables in comparison with structural modeling.

approach, because it models agent behavior, specifies all the structural relationships in the agent's optimization problem. Empirical implementation of the behavioral model requires that the variables that enter the structural components be measured. The PROGRESA program extensively surveyed the families in both the treatment and control villages, which allowed researchers to specify and estimate rich models of family behavior.

As illustrated by the papers based on the PROGRESA data, researchers do not always choose the same behavioral model. That researchers disagree a priori on the appropriate behavioral model leads to an issue that is probably the most vexing in empirical research, namely that of model validation and selection. Here too, as illustrated by several of the papers, there is a synergy between field experiments and structural estimation. Having a controlled experiment provides a natural holdout sample with which to perform an external model validation (Todd & Wolpin 2006, Duflo et al. 2008). Such a tool is particularly important in the case of structural estimation because the ability of the DCDP approach to perform counterfactuals is offset by the reliance of DCDP models on extratheoretic modeling choices, including functional form and distributional assumptions. As illustrated by the Kaboski & Townsend (2007) paper, however, having a field experiment is not necessary for external validation, as holdout samples can be strategically chosen from observational data as well.[24]

Although the first DCDP model (Rosenzweig & Wolpin 1993) applied to development economics appeared over 15 years ago, the adoption of this approach in development economics has lagged behind that of other applied areas of economics. The recent recognition of its value, particularly when combined with field experiments, will likely spur further applications.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Chuck Manski for comments.

LITERATURE CITED

Aguirregabiria V, Mira P. 2010. Dynamic discrete choice structural models: a survey. *J. Econom.* In press

Aiyagari SR. 1994. Uninsured idiosyncratic risk and aggregate saving. *Q. J. Econ.* 109:659–84

Attanasio O, Meghir C, Santiago A. 2005. *Education choices in Mexico: using a structural model and a randomized experiment to evaluate Progresa*. Work. Pap., University College, London

Behrman JR, Sengupta P, Todd PE. 2005. Progressing through PROGRESA: an impact assessment of a school subsidy experiment in rural Mexico. *Econ. Dev. Cult. Change* 54(1):237–75

Bravo D, Mukhopadhyay S, Todd PE. 2009. *Effects of a universal school voucher system on educational and labor market outcomes: evidence from Chile*. Work. Pap., Univ. Pennsylvania

[24]The use of holdout samples, either based on randomized experiments or observational data, is not new. Examples outside of the development literature include Wise (1985), Lumsdaine et al. (1994), and Keane & Wolpin (2007).

Colussi A. 2006. *Migrants' networks: an estimable model of illegal Mexican migration*. Work. Pap., Univ. Western Ontario, Canada

Deaton A. 1991. Saving and liquidity constraints. *Econometrica* 59(5):1221–48

Deaton A. 2009. *Instruments of development: randomization in the Tropics and the search for the elusive keys to economic development*. NBER Work. Pap. 14690

Dominitz J, Manski CF. 1996. Eliciting student expectations of the returns to schooling. *J. Hum. Resour.* 31(1)1–26

Dominitz J, Manski CF. 1997. Using expectations data to study subjective income expectations. *J. Am. Stat. Assoc.* 92:855–67

Duflo E, Hanna E, Ryan S. 2008. *Monitoring works: getting teachers to come to school*. Work. Pap., MIT

Eckstein Z, Wolpin KI. 1989. The specification and estimation of dynamic stochastic discrete choice models: a survey. *J. Hum. Resour.* 24(4):562–98

Friedman M. 1955. The role of government in education. In *Economics and the Public Interest*, ed. RA Colo, pp. 123–44. New Jersey: Rutgers Univ. Press

Friedman M. 1962. *Capitalism and Freedom*. Princeton, NJ: Princeton Univ. Press

Gotz GA, McCall JJ. 1984. *Estimating the Stay/Leave Decision in a Sequential Decision-Making Model*. Santa Monica, CA: Rand Corp.

Heckman JJ. 1979. Sample selection bias as a specification error. *Econometrica* 47:153–62

Heckman JJ. 1981. The incidental parameters problem of initial conditions in estimating a discrete time-discrete data stochastic process and some Monte Carlo evidence. In *Structural Analysis of Discrete Data with Econometric Applications*, ed. CF Manski, D McFadden, pp. 179–97. Cambridge, MA: MIT Press

Heckman JJ, Urzua S. 2009. *Comparing IV with structural models: What simple IV can and cannot identify.* NBER Work. Pap. 14706

Kaboski JP, Townsend RM. 2007. *Testing a structural model of credit constraints using a large-scale quasi-experimental microfinance initiative*. Work. Pap., MIT

Keane M, Wolpin KI. 2007. Exploring the usefulness of a non-random holdout sample for model validation: welfare effects on female behavior. *Int. Econ. Rev.* 48:1351–78

Keane M, Wolpin KI. 2009. Empirical applications of discrete choice dynamic programming models. *Rev. Econ. Dyn.* 1:1–22

Lumsdaine RL, Stock JH, Wise DA. 1994. Pension plan provisions and retirement: men and women, medicare, and models. In *Studies in the Economics of Aging*, ed. DA Wise. 2004. Measuring expectations. *Econometrica* 72(5):1329–76

McKee D. 2006. *A dynamic model of retirement in Indonesia*. PhD diss., Univ. Calif., Los Angeles

Miller RA. 1984. Job matching and occupational choice. *J. Polit. Econ.* 92:1086–120

Miller RA. 1997. Estimating models of dynamic optimization with microeconomic data. In *Handbook of Applied Econometrics: Microeconometrics*, ed. MH Pesaran, P Schmidt, pp. 247–99. Oxford: Blackwell

Pakes A. 1986. Patents as options: some estimates of the value of holding European patent stocks. *Econometrica* 54(4):755–84

Rosenzweig MR, Wolpin KI. 1993. Credit market constraints, consumption smoothing, and the accumulation of durable production assets. *J. Polit. Econ.* 101(2):223–44

Rust J. 1987. Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher. *Econometrica* 55(5)999–1034

Rust J. 1994a. Numerical dynamic programming in economics. In *Handbook of Computational Economics*, ed. K Amman, J Rust, pp. 619–729. Amsterdam: North Holland

Rust J. 1994b. Structural estimation of Markov decision models. In *Handbook of Econometrics*, Vol. 4, ed. R Engle, D McFadden , pp. 3082–143. Amsterdam: North Holland

Schultz TP. 2004. School subsidies for the poor: evaluating the Mexican Progresa poverty program. *J. Dev. Econ.* 74(1):199–250

Todd PE, Wolpin KI. 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. *Am. Econ. Rev.* 96(5):1384–417

Todd PE, Wolpin KI. 2010. Ex-ante evaluation of social programs. *Ann. Econ. Stat.* In press

Van der Klaauw W. 2000. *On the use of expectations data in estimating structural dynamic models: an analysis of career choices*. Work. Pap., Univ. North Carolina, Chapel Hill

Vélez-Grajales V. 2009. *Reforms to an individual account pension system and their effects on work and contribution decisions: the case of Chile*. PhD diss., Univ. Pennsylvania, Philadelphia

Wise DA. 1985. Behavioral model verses experimentaion: the effects of housing subsidies on rent. In *Methods of Operations Research*, Vol. 50, ed. P Brucker, R Pauly, pp. 441–89. Konigstein: Verlag Anton Hain

Wolpin KI. 1984. A dynamic stochastic model of fertility and child mortality. *J. Polit. Econ.* 92:852–74

Wolpin KI. 1987. Estimating a structural search model: the transition from school to work. *Econometrica* 55(4):801–17