Ex Ante Policy Evaluation

Kenneth I. Wolpin
University of Pennsylvania

Koopmans Memorial Lecture (1)
Cowles Foundation
Yale University
November 3, 2010

Introduction

Goals of evaluation:

- *Ex post* evaluation – evaluate an existing policy subsequent to its implementation - a mainstay of social science research.

Introduction

Goals of evaluation:

- *Ex post* evaluation – evaluate an existing policy subsequent to its implementation - a mainstay of social science research.

- *Ex ante* evaluation – evaluate policies that are outside of the historical experience, e.g., completely new programs, adding features to existing programs.

Introduction

Evaluation methodologies:

- Experimental - evaluation of pilot or demonstration project involving treatment randomization.

Introduction

Evaluation methodologies:

- Experimental - evaluation of pilot or demonstration project involving treatment randomization.

- Non-experimental – evaluation of observational data using a combination of behavioral and statistical assumptions.

Introduction

Purpose of this talk:

- Discuss the development of alternative non-experimental methods for *ex ante* policy evaluation
  - Nonstructural – Nonparametric
  - Structural – Parametric

Introduction

Purpose of this talk:

- Discuss the development of alternative non-experimental methods for *ex ante* policy evaluation
  - Nonstructural – Nonparametric
  - Structural – Parametric

- Present an application that combines experimental data and these non-experimental methods.

Introduction

Purpose of this talk:

- Discuss the development of alternative non-experimental methods for *ex ante* policy evaluation
  - Nonstructural – Nonparametric
  - Structural – Parametric

- Present an application that combines experimental data and these non-experimental methods.

- Use the application to explore model validation and model selection issues.

Nonstructural - Nonparametric Approach

- Roots in Marschak (1953)

- Methodology recently revisited
  - Heckman (2000), Ichimura and Taber (2000), Todd and Wolpin (2008)

- Recent empirical applications
  - Ichimura and Taber (2000),Blomquist and Newey (2002),Todd and Wolpin (2008), Azevedo, Bouillon and Yanez-Pagans (2009),Thomas (2010), Zantamo, Pudney and Hancock (2010).

Working Example

Consider the problem of evaluating the impact of a new policy proposal to provide a monetary subsidy to low-income households based on the school attendance of their children.

Working Example

Consider the problem of evaluating the impact of a new policy proposal to provide a monetary subsidy to low-income households based on the school attendance of their children.

Assume that in the current setting, there is no direct tuition cost of schooling. Tuition variation cannot be used to extrapolate the effect of the subsidy (negative tuition).

NS-NP Approach

One way to approach the evaluation problem is to specify a model of household behavior.

1. This discussion is based on Todd and Wolpin (Annales D'Economie et de Statistique, 2008)

NS-NP Approach

One way to approach the evaluation problem is to specify a model of household behavior.

Assume a household has one child and solves the following optimization problem in the absence of the intervention:

$$\max_{\{s\}} U(c, s; x, \epsilon)$$

subject to $c = y + w(1 - s)$

where $s = 1$ if the child attends school and $s = 0$ if the child works for pay.

NS-NP Approach

$$\max_{\{s\}} U(c, s; x, \epsilon)$$

subject to $c = y + w(1 - s)$

$$\Rightarrow \quad s^* = \varphi(y, w; x, \epsilon)$$

NS-NP Approach

Introducing an attendance subsidy of τ , the household faces the new budget constraint:

$$c = y + w(1 - s) + \tau s$$

NS-NP Approach

Introducing an attendance subsidy of τ , the household faces the new budget constraint:

$$c = y + w(1 - s) + \tau s$$

Rewriting,

$$c = (y + \tau) + (w - \tau)(1 - s)$$

NS-NP Approach

Introducing an attendance subsidy of τ , the household faces the new budget constraint:

$$c = y + w(1-s) + \tau s$$

Rewriting,

$$\begin{aligned} c &= (y+\tau) + (w-\tau)(1-s) \\ &= \widetilde{y} + \widetilde{w}(1-s) \end{aligned}$$

NS-NP Approach

Introducing an attendance subsidy of τ , the household faces the new budget constraint:

$$c = y + w(1-s) + \tau s$$

Rewriting,

$$\begin{aligned} c &= (y+\tau) + (w-\tau)(1-s) \\ &= \widetilde{y} + \widetilde{w}(1-s) \end{aligned}$$

The optimal choice in the presence of the subsidy is

$$s^* = \varphi(\widetilde{y}, \widetilde{w}; x, \epsilon)$$

NS-NP Approach

A household characterized by the vector

$$\{y, w, x, \epsilon\}$$

would make the same school attendance decision in the presence of the subsidy as a household characterized by the vector

$$\{\widetilde{y} = y + \tau, \widetilde{w} = w - \tau, x, \epsilon\}$$

would make without the subsidy.

NS-NP Approach

Under the assumption that

$$f(\epsilon|y, w, x) = f(\epsilon|\widetilde{y}, \widetilde{w}, x)$$

and given wage data for *non-working* children, a consistent estimator of the effect of the subsidy program on school attendance is

$$\frac{1}{N} \sum_{\substack{j=1 \\ j,i \in S_j \times S_i}}^{N} \left\{ \widehat{E}(s_i | y_i = \widetilde{y}_j, w_i = \widetilde{w}_j, x_i = x_j) - s_j(y_j, w_j, x_j) \right\}$$

This estimator can be implemented non-parametrically with a matching procedure (as will be shown below).

NS-NP Approach

Extensions:

1. Multiple children - exogenous and endogenous fertility.

NS-NP Approach

Extensions:

1. Multiple children - exogenous and endogenous fertility.

2. Multiple periods with perfect foresight.

NS-NP Approach

Extensions:

1. Multiple children - exogenous and endogenous fertility.

2. Multiple periods with perfect foresight.

3. Partial observability of child wages (requires a distributional assumption for wages, but not for preferences).

NS-NP Approach

Limitations:

1. Cannot match on unobservables – must maintain independence assumption (conditional on observables) $\Rightarrow$ generally not applicable to dynamic models.

NS-NP Approach

Limitations:

1. Cannot match on unobservables – must maintain independence assumption (conditional on observables) $\Rightarrow$ generally not applicable to dynamic models.

2. Curse of dimensionality

NS-NP Approach

Limitations:

1. Cannot match on unobservables – must maintain independence assumption (conditional on observables) $\Rightarrow$ generally not applicable to dynamic models.

2. Curse of dimensionality

3. Requires large samples

NS-NP Approach

Limitations:

1. Cannot match on unobservables – must maintain independence assumption (conditional on observables) $\Rightarrow$ generally not applicable to dynamic models.

2. Curse of dimensionality

3. Requires large samples

4. Potential ambiguity in model x policy space.

NS-NP Approach

Suppose we add a value of child home production,

$$\max_{\{s,l\}} U(c, s, l; x, \epsilon_s, \epsilon_l)$$

where $(s, l) \in \{(0,0), (0,1), (1,0)\}$

NS-NP Approach

Suppose we add a value of child home production,

$$\max_{\{s,l\}} U(c, s, l; x, \epsilon_s, \epsilon_l)$$

where $(s, l) \in \{(0,0), (0,1), (1,0)\}$

subject to the budget constraint

$$c = y + w(1 - s - l)$$

NS-NP Approach

Suppose we add a value of child home production,

$$\max_{\{s,l\}} U(c, s, l; x, \epsilon_s, \epsilon_l)$$

where $(s, l) \in \{(0,0), (0,1), (1,0)\}$

subject to the budget constraint

$$c = y + w(1 - s - l)$$

$$\Rightarrow \qquad s^* = \psi^s(y, w; x, \epsilon_s, \epsilon_l)$$

NS-NP Approach

With a school subsidy, the budget constraint is

$$c = y + w(1 - s - l) + \tau s$$

NS-NP Approach

With a school subsidy, the budget constraint is

$$c = y + w(1 - s - l) + \tau s$$

$$= (y + \tau) + (w - \tau)(1 - s - l) - \tau l$$

NS-NP Approach

With a school subsidy, the budget constraint is

$$c = y + w(1 - s - l) + \tau s$$

$$= (y + \tau) + (w - \tau)(1 - s - l) - \tau l$$

$$= \widetilde{y} + \widetilde{w}(1 - s - l) - \tau l$$

NS-NP Approach

With a school subsidy, the budget constraint is

$$c = y + w(1 - s - l) + \tau s$$

$$= (y + \tau) + (w - \tau)(1 - s - l) - \tau l$$

$$= \widetilde{y} + \widetilde{w}(1 - s - l) - \tau l$$

$\Rightarrow$

$$s^* \neq \psi^s(\widetilde{y}, \widetilde{w}; x, \epsilon_s, \epsilon_l)$$

NS-NP Approach

With a school subsidy, the budget constraint is

$$c = y + w(1 - s - l) + \tau s$$

$$= (y + \tau) + (w - \tau)(1 - s - l) - \tau l$$

$$= \widetilde{y} + \widetilde{w}(1 - s - l) - \tau l$$

$\Rightarrow$

$$s^* \neq \psi^s(\widetilde{y}, \widetilde{w}; x, \epsilon_s, \epsilon_l)$$

s^* depends also on τ .

NS-NP Approach

Suppose instead that the policy provides a subsidy to the household as long as the child does no market work.

The budget constraint is then

$$c = y + w(1 - s - l) + \tau(s + l)$$

NS-NP Approach

Suppose instead that the policy provides a subsidy to the household as long as the child does no market work.

The budget constraint is then

$$c = y + w(1 - s - l) + \tau(s + l)$$

$$= (y + \tau) + (w - \tau)(1 - s - l)$$

NS-NP Approach

Suppose instead that the policy provides a subsidy to the household as long as the child does no market work.

The budget constraint is then

$$c = y + w(1 - s - l) + \tau(s + l)$$

$$= (y + \tau) + (w - \tau)(1 - s - l)$$

$$= \widetilde{y} + \widetilde{w}(1 - s - l)$$

NS-NP Approach

Suppose instead that the policy provides a subsidy to the household as long as the child does no market work.

The budget constraint is then

$$c = y + w(1 - s - l) + \tau(s + l)$$

$$= (y + \tau) + (w - \tau)(1 - s - l)$$

$$= \widetilde{y} + \widetilde{w}(1 - s - l)$$

$\Rightarrow$

$$s^* = \psi^s(\widetilde{y}, \widetilde{w}; x, \epsilon_s, \epsilon_l)$$

NS-NP Approach

The estimator based on a comparison of the behavior of households with $\{y, w\}$ to those with $\{\widetilde{y} = y + \tau, \widetilde{w} = w - \tau\}$ is consistent with either of two models and policies:

NS-NP Approach

The estimator based on a comparison of the behavior of households with $\{y, w\}$ to those with $\{\widetilde{y} = y + \tau, \widetilde{w} = w - \tau\}$ is consistent with either of two models and policies:

$U(c, s; x, \epsilon)$ and subsidy if attend school

NS-NP Approach

The estimator based on a comparison of the behavior of households with $\{y, w\}$ to those with $\{\widetilde{y} = y + \tau, \widetilde{w} = w - \tau\}$ is consistent with either of two models and policies:

$U(c, s; x, \epsilon)$ and subsidy if attend school

or

$U(c, s, l; x, \epsilon)$ and subsidy if not work

NS-NP Approach

The estimator based on a comparison of the behavior of households with $\{y, w\}$ to those with $\{\widetilde{y} = y + \tau, \widetilde{w} = w - \tau\}$ is consistent with either of two models and policies:

$U(c, s; x, \epsilon)$ and subsidy if attend school

or

$U(c, s, l; x, \epsilon)$ and subsidy if not work

It is necessary to take a stand on the arguments of the utility function (i.e., on the model).

Structural - Parametric Approach
Discrete Choice Dynamic Programming Models

Structural - Parametric Approach
Discrete Choice Dynamic Programming Models

The development of methods for the estimation of discrete choice dynamic programming (DCDP) models, which began 25 years ago, opened up new frontiers for empirical research in a number of areas:

- labor economics
- industrial organization
- economic demography
- development economics
- health economics
- political economy
- law and economics

S-P Approach

The literature began with independent contributions by

- Gotz and McCall (1984, unpub.)
- Miller (1984, JPE)
- Pakes (1986, EMA)
- Rust (1987, EMA)
- Wolpin (1984, JPE; 1987, EMA)

S-P Approach

Basic insight for the estimation of DCDP models:

DCDP models can be cast as static estimation problems.

S-P Approach

Static Model:

Parents in household i decide on whether to send their (only) child to school $s_{it} = 1$ or to work $s_{it} = 0$.

S-P Approach

Static Model:

Parents in household i decide on whether to send their (only) child to school $s_{it} = 1$ or to work $s_{it} = 0$.

In each period, parents choose s_{it} to maximize

$$U_{it} = C_{it} + \alpha_{it} s_{it}$$

where

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

S-P Approach

Static Model:

Parents in household i decide on whether to send their (only) child to school $s_{it} = 1$ or to work $s_{it} = 0$.

In each period, parents choose s_{it} to maximize

$$U_{it} = C_{it} + \alpha_{it} s_{it}$$

where

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

subject to the budget constraint

$$C_{it} = y_{it} + w_{it}(1 - s_{it})$$

S-P Approach

Define the alternative-specific utilities:

$$U_{it}^{1} = y_{it} + x_{it}\beta + \epsilon_{it}$$

$$U_{it}^{0} = y_{it} + w_{it}$$

S-P Approach

Define the alternative-specific utilities:

$$U_{it}^{1} = y_{it} + x_{it}\beta + \epsilon_{it}$$

$$U_{it}^{0} = y_{it} + w_{it}$$

The difference in alternative specific utilities (the latent variable function) that governs the choice is:

$$U_{it}^{1} - U_{it}^{0} = x_{it}\beta - w_{it} + \epsilon_{it}$$

S-P Approach

Let

$\Omega_{it} = \{x_{it}, w_{it}, \epsilon_{it}\}$ be the household's state space at t,

$\Omega_{it}^{-} = \{x_{it}, w_{it}\}$ be the part of the household's state space observed by the researcher.

S-P Approach

Let

$\Omega_{it} = \{x_{it}, w_{it}, \epsilon_{it}\}$ be the household's state space at t,
$\Omega_{it}^{-} = \{x_{it}, w_{it}\}$ be the part of the household's state space observed by the researcher.

The value of the preference unobservable ϵ_{it} that makes the parents indifferent between sending the child to school or to work is

$$\epsilon_{it}^{*}(\Omega_{it}^{-}) = w_{it} - x_{it}\beta$$

S-P Approach

Let

$\Omega_{it} = \{x_{it}, w_{it}, \epsilon_{it}\}$ be the household's state space at t,
$\Omega_{it}^{-} = \{x_{it}, w_{it}\}$ be the part of the household's state space observed by the researcher.

The value of the preference unobservable ϵ_{it} that makes the parents indifferent between sending the child to school or to work is

$$\epsilon_{it}^{*}(\Omega_{it}^{-}) = w_{it} - x_{it}\beta$$

so that

$$U_{it}^{1} - U_{it}^{0} = -\epsilon_{it}^{*}(\Omega_{it}^{-}) + \epsilon_{it}$$

S-P Approach

Cross-Section Data: $(s_{it}, x_{it}, w_{it} : i = 1, ..., I)$

S-P Approach

Cross-Section Data: $(s_{it}, x_{it}, w_{it} : i = 1, ..., I)$

Suppose $\epsilon_{it} \backsim N(0, \sigma_\epsilon)$ and independent of the elements of Ω_{it}^-.

S-P Approach

Cross-Section Data: $(s_{it}, x_{it}, w_{it} : i = 1, ..., I)$

Suppose $\epsilon_{it} \backsim N(0, \sigma_\epsilon)$ and independent of the elements of Ω_{it}^{-}.

The likelihood function is:

$$
\begin{aligned}
&\pounds(\beta, \sigma_\epsilon | s_{it}, x_{it}, w_{it}) \\
&= \prod_{i=1}^{i=I} \Pr(\epsilon_{it} > \epsilon_{it}^{*}(\Omega_{it}^{-}))^{s_{it}} (1 - \Pr(\epsilon_{it} > \epsilon_{it}^{*}(\Omega_{it}^{-}))^{1-s_{it}} \\
&= \prod_{i=1}^{i=I} \Phi\left(\frac{-w_{it} + x_{it}\beta}{\sigma_\epsilon}\right)^{s_{it}} \left(1 - \Phi\left(\frac{-w_{it} + x_{it}\beta}{\sigma_\epsilon}\right)\right)^{1-s_{it}}
\end{aligned}
$$

S-P Approach

σ_ϵ is identified from wage variation (assumed to be observed for both children who work and children who do not) and, given that, β is identified from variation in x's.

S-P Approach

σ_ϵ is identified from wage variation (assumed to be observed for both children who work and children who do not) and, given that, β is identified from variation in x's.

It is rare that wages are observed for non-workers. In that case, one must specify how wage offers are determined.

S-P Approach

σ_ϵ is identified from wage variation (assumed to be observed for both children who work and children who do not) and, given that, β is identified from variation in x's.

It is rare that wages are observed for non-workers. In that case, one must specify how wage offers are determined.

Assume $w_{it} = z_{it}\gamma + \eta_{it}$
where $f(\epsilon_{it}, \eta_{it}) \backsim N(0, \Lambda)$ and independent of the elements of Ω_{it}^{-} (which now includes z_{it}). The errors are mutually serially independent.

S-P Approach

Write the latent variable function as:

$$
\begin{aligned}
U_{it}^1 - U_{it}^0 &= x_{it}\beta - z_{it}\gamma + \epsilon_{it} - \eta_{it} \\
&= -\xi_{it}^*(\Omega_{it}^-) + \xi_{it}
\end{aligned}
$$

S-P Approach

Write the latent variable function as:

$$\begin{aligned} U_{it}^1 - U_{it}^0 &= x_{it}\beta - z_{it}\gamma + \epsilon_{it} - \eta_{it} \\ &= -\xi_{it}^*(\Omega_{it}^-) + \xi_{it} \end{aligned}$$

Thus,

$$\Pr(s_{it} = 1) = \Phi\left(\frac{-x_{it}\beta + z_{it}\gamma}{\sigma_\xi}\right)$$

S-P Approach

Write the latent variable function as:

$$\begin{aligned} U_{it}^1 - U_{it}^0 &= x_{it}\beta - z_{it}\gamma + \epsilon_{it} - \eta_{it} \\ &= -\xi_{it}^*(\Omega_{it}^-) + \xi_{it} \end{aligned}$$

Thus,

$$\Pr(s_{it} = 1) = \Phi\left(\frac{-x_{it}\beta + z_{it}\gamma}{\sigma_\xi}\right)$$

With partial observability, as in the case in which all wage offers are observed, identification requires that there be independent variation in wages, in this case through z_{it}.

S-P Approach

$$\Pr(s_{it} = 1) = \Phi\left(\frac{-x_{it}\beta + z_{it}\gamma}{\sigma_\xi}\right)$$

Identification requires at least one variable that affects the wage offer, a variable in z_{it}, that does not affect the utility from attending school, a variable in x_{it}.

S-P Approach

Dynamic Model:

There are a number of ways to introduce dynamics in the model – in preferences, in constraints, e.g.,

S-P Approach

Dynamic Model:

There are a number of ways to introduce dynamics in the model – in preferences, in constraints, e.g.,

(1) Parents may care about their child's school attainment at some terminal age - attending school in any period affects future utility:

$$U_{iT} = \kappa S_{iT}$$

$$S_{i,t+1} = \sum_{\tau=1}^{t} s_{i\tau} = S_{it} + s_{it}$$

S-P Approach

(2) The wage a child earns might depend on the child's work experience – working in any period affects future wages:

$$w_{it} = z_{it}\gamma_1 + \gamma_2 H_{it} + \eta_{it}$$

$$H_{i,t+1} = \sum_{\tau=1}^{t} (1 - s_{i\tau}) = H_{it} + (1 - s_{it})$$

S-P Approach

Value functions, etc.

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

S-P Approach

Value functions, etc.

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

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

S-P Approach

Value functions, etc.

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

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

$$\begin{aligned} V_t^k(\Omega_{it}) &= U^k_{it}(\Omega_{it}) + \delta E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}, s_{it} = k] \quad t < T, \\ &= U^k_{iT}(\Omega_{iT}) \quad t = T. \end{aligned}$$

S-P Approach

Latent Variable Function

$$V_t^1(\Omega_{it}) - V_t^0(\Omega_{it})$$

$$= \quad \underbrace{x_{it}\beta - z_{it}\gamma_1}_{-\xi^*_{it}(\Omega^-_{it})} - \gamma_2 H_{it} + \underbrace{\epsilon_{it} - \eta_{it}}_{+\,\xi_{it}}$$

S-P Approach

Latent Variable Function

$$V_t^1(\Omega_{it}) - V_t^0(\Omega_{it})$$

$$= \quad x_{it}\underbrace{\beta - z_{it}}_{-\xi^*_{it}(\Omega^-_{it})}\gamma_1 - \gamma_2 H_{it} + \underbrace{\epsilon_{it} - \eta_{it}}_{+\,\xi_{it}}$$

$$+\delta\left\{[E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}, s_{it} = 1] - [E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}, s_{it} = 0]\right\}$$

S-P Approach

Latent Variable Function

$$V_t^1(\Omega_{it}) - V_t^0(\Omega_{it})$$

$$= \quad x_{it}\underbrace{\beta - z_{it}}_{-\xi^*_{it}(\Omega^-_{it})}\gamma_1 - \gamma_2 H_{it} + \underbrace{\epsilon_{it} - \eta_{it}}_{+\,\xi_{it}}$$

$$+\delta\left\{[E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}, s_{it} = 1] - [E[V_{t+1}(\Omega_{i,t+1})|\Omega_{it}, s_{it} = 0]\right\}$$

$$= \quad -\xi^{**}_{it}(\Omega^-_{it}) + \xi_{it}$$

S-P Approach

Static case:

$$\Pr(s_{it} = 1) = \Pr(\xi_{it} > \xi^*_{it}(x_{it}, z_{it}; \beta, \gamma_1, \Lambda))$$

S-P Approach

Static case:

$$\Pr(s_{it} = 1) = \Pr(\xi_{it} > \xi^{*}_{it}(x_{it}, z_{it}; \beta, \gamma_1, \Lambda))$$

Dynamic case:

$$\Pr(s_{it} = 1) = \Pr(\xi_{it} > \xi^{**}_{it}(x_{it}, z_{it}, H_{it}; \beta, \gamma_1, \Lambda, \gamma_2, \delta))$$

S-P Approach

The effect of an attendance subsidy, τ , on the probability that a child attends school is

$$\Phi\left(\frac{-\xi_{it}^{**}(\Omega_{it}^{-})+\tau}{\sigma_{\xi}}\right)-\Phi\left(\frac{-\xi_{it}^{**}(\Omega_{it}^{-})}{\sigma_{\xi}}\right)$$

S-P Approach

The effect of an attendance subsidy, τ , on the probability that a child attends school is

$$\Phi\left(\frac{-\xi_{it}^{**}(\Omega_{it}^{-})+\tau}{\sigma_{\xi}}\right)-\Phi\left(\frac{-\xi_{it}^{**}(\Omega_{it}^{-})}{\sigma_{\xi}}\right)$$

Identification of the subsidy effect requires knowledge of σ_{ξ} .

S-P Approach

The effect of an attendance subsidy, τ , on the probability that a child attends school is

$$\Phi\left(\frac{-\xi_{it}^{**}(\Omega_{it}^{-})+\tau}{\sigma_\xi}\right) - \Phi\left(\frac{-\xi_{it}^{**}(\Omega_{it}^{-})}{\sigma_\xi}\right)$$

Identification of the subsidy effect requires knowledge of σ_ξ .

The same exclusion restriction as for identification in the static model, that there be a variable in the wage function that does not affect preferences, is also necessary in the dynamic model.

Extensions and Advances in DCDP Modeling

1. Multinomial Choice - With N dichotomous choice variables, there are at most $K = 2^N$ possible mutually exclusive choices, K alternative-specific value functions and K -1 latent variable functions.

Extensions and Advances in DCDP Modeling

1. Multinomial Choice - With N dichotomous choice variables, there are at most $K = 2^N$ possible mutually exclusive choices, K alternative-specific value functions and K -1 latent variable functions.

   Closed form solutions for DCDP models and likelihood function using specific functional forms and distributional assumptions (Rust,1987)

Extensions and Advances in DCDP Modeling

1. Multinomial Choice - With N dichotomous choice variables, there are at most $K = 2^N$ possible mutually exclusive choices, K alternative-specific value functions and K -1 latent variable functions.

   Closed form solutions for DCDP models and likelihood function using specific functional forms and distributional assumptions (Rust,1987)

   Value function approximation and simulation methods of estimation coupled with increases in computational speed have enabled the estimation of models with large choice sets and state spaces (Keane and Wolpin, 1994).

Extensions and Advances in DCDP Modeling

2. Unobserved Heterogeneity

   Allows for permanent differences in preferences and constraints among agents. A common specification is to allow for a fixed number of agent types.

Extensions and Advances in DCDP Modeling

2. Unobserved Heterogeneity

   Allows for permanent differences in preferences and constraints among agents. A common specification is to allow for a fixed number of agent types.

3. Flexible specifications

   Any DCDP that can be numerically solved can, in principle, be estimated accommodating:

   Non-additive errors

   Serial correlation in unobservables

   Alternative functional forms

   Alternative distributional assumptions

Extensions and Advances in DCDP Modeling

4. Relaxation of parametric assumptions: Heckman and Navarro (2007).

Extensions and Advances in DCDP Modeling

4. Relaxation of parametric assumptions: Heckman and Navarro (2007).

5. Alternative estimation approaches

   Non-full solution methods: Hotz and Miller (1993, 1994), Arcidiacono and Miller (2006).

   Bayesian methods: Imai, Jain and Ching (2009), Norets (2009).

Neither the structural-parametric nor the nonstructural-nonparametric approach to *ex ante* policy evaluation is assumption free.

Neither the structural-parametric nor the nonstructural-nonparametric approach to *ex ante* policy evaluation is assumption free.

A common element in the two approaches is the necessity for specifying a theory.

Application - PROGRESA

In 1997, Mexico initiated the PROGRESA program to increase schooling levels of children in rural areas.

Application - PROGRESA

In 1997, Mexico initiated the PROGRESA program to increase schooling levels of children in rural areas.

The CCT (conditional cash transfer) program provided a subsidy to low-income families for sending their children to school.

Application - PROGRESA

In 1997, Mexico initiated the PROGRESA program to increase schooling levels of children in rural areas.

The CCT (conditional cash transfer) program provided a subsidy to low-income families for sending their children to school.

The initial program has been extended to urban areas in Mexico (and renamed Oportunidades), and adopted in numerous other countries (for example, Bangladesh, Brazil, Colombia, Nicaragua, Pakistan).

Table 4
Monthly Transfers for School Attendance under the PROGRESA Program

School Level	Grade	Monthly Payment in Pesos	
		Females	Males
Primary	3	70	70
	4	80	80
	5	105	105
	6	135	135
Secondary	1	210	200
	2	235	210
	3	255	225

Source: Schultz (1999a, Table 1).
Corresponds to first term of the 1998-99 school year.

Application - PROGRESA

To evaluate the program, the Mexican government conducted a randomized social experiment.

Application - PROGRESA

To evaluate the program, the Mexican government conducted a randomized social experiment.

- 506 rural villages were randomly assigned to either participate in the program or serve as controls.

Application - PROGRESA

To evaluate the program, the Mexican government conducted a randomized social experiment.

- 506 rural villages were randomly assigned to either participate in the program or serve as controls.

- As part of the evaluation, baseline and follow-up surveys were conducted obtaining detailed information on household demographics, parental income and the school attendance, work and earnings of children.

NS-NP Implementation

Basic features of the models:

NS-NP:

$U(c, s_1, s_2, ..., s_n; x, \epsilon)$; x is a vector of child ages and genders.

$$c = (y + \sum_{j=1}^{n} \tau(g_j, S_j)) + \sum_{j=1}^{n} (w_j - \tau(g_j, S_j))(1 - s_j)$$

where g_j is child j's gender and S_j is the child j's schooling level.

NS-NP Implementation

Basic features of the models:

NS-NP:

$U(c, s_1, s_2, ..., s_n; x, \epsilon)$; x is a vector of child ages and genders.

$$c = (y + \sum_{j=1}^{n} \tau(g_j, S_j)) + \sum_{j=1}^{n} (w_j - \tau(g_j, S_j))(1 - s_j)$$

where g_j is child j’s gender and S_j is the child j’s schooling level.

The child wage is a reported *village level* minimum wage. Matches are to different villages.

NS-NP Implementation

Estimate with two different sets of covariates:

1. Match only on child age and gender (single child);

NS-NP Implementation

Estimate with two different sets of covariates:

1. Match only on child age and gender (single child);
2. Match on child age, gender and the number of children in the household (multiple child).

NS-NP Implementation

Estimate with two different sets of covariates:

1. Match only on child age and gender (single child);
2. Match on child age, gender and the number of children in the household (multiple child).

The model is estimated on control households post-program.

NS-NP Implementation

Single child specification:

We use a two-dimensional kernel regression estimator:

$$\widehat{E}(s_i | y_i = \widetilde{y}_j, w_i = \widetilde{w}_j, \chi_i = \chi_j) = \frac{\sum\limits_{\substack{i=1 \\ i \in S_p}}^{N} s_i K\left(\frac{w_i - \widetilde{w}_j}{h_n^w}\right) K\left(\frac{y_i - \widetilde{y}_j}{h_n^y}\right) 1(\chi_i = \chi_j)}{\sum\limits_{\substack{i=1 \\ i \in S_p}}^{N} K\left(\frac{w_i - \widetilde{w}_j}{h_n^w}\right) K\left(\frac{y_i - \widetilde{y}_j}{h_n^y}\right) 1(\chi_i = \chi_j)}$$

where $K(\cdot)$ denotes the kernel function, h_n^w and h_n^y are bandwidth parameters and χ the matched covariates.

NS-NP Implementation

Single child specification:

We use a two-dimensional kernel regression estimator:

$$\widehat{E}(s_i|y_i = \widetilde{y}_j, w_i = \widetilde{w}_j, \chi_i = \chi_j) = \frac{\sum\limits_{\substack{i=1 \\ i \in S_p}}^{N} s_i K\left(\frac{w_i - \widetilde{w}_j}{h_n^w}\right) K\left(\frac{y_i - \widetilde{y}_j}{h_n^y}\right) 1(\chi_i = \chi_j)}{\sum\limits_{\substack{i=1 \\ i \in S_p}}^{N} K\left(\frac{w_i - \widetilde{w}_j}{h_n^w}\right) K\left(\frac{y_i - \widetilde{y}_j}{h_n^y}\right) 1(\chi_i = \chi_j)}$$

where $K(\cdot)$ denotes the kernel function, h_n^w and h_n^y are bandwidth parameters and χ the matched covariates.

$$\widehat{\alpha} = \frac{1}{N} \sum_{\substack{j=1 \\ j, i \in S_P}}^{N} \left\{ \widehat{E}(s_i|y_i = \widetilde{y}_j, w_i = \widetilde{w}_j, \chi_i = \chi_j) - s_j(y_j, w_j, \chi_j) \right\}$$

NS-NP Implementation

Multiple children specification:

$$\widehat{\alpha} = \frac{1}{N} \sum_{\substack{j=1 \\ j,i \in S_P}}^{N} \left\{ \widehat{E}(s_i | y_i = y_j + n_j \overline{\tau}_j, w_i = w_j - \overline{\tau}_j, \chi_i = \chi_j) - s_j(y_j, w_j, \chi_j) \right\}$$

where $\overline{\tau}_j$ is the average subsidy for children in the family and where all children in the family face the same (village level) wage.

S-P Implementation

Basic features

DCDP model:

Choice set: school attendance of all children in household ages 6-15, employment of all children ages 12-15, fertility.

S-P Implementation

DCDP model:

Choice set: school attendance of all children in household ages 6-15, employment of all children ages 12-15, fertility.

Utility: gender-specific school attendance and attainment, gender- and age-specific home value that depends on whether younger children are at home, nonseparable in consumption and schooling, heterogeneous in unobserved types.

S-P Implementation

Child wage offer function: gender, age, unobserved productivity type, distance to nearest city.

S-P Implementation

Child wage offer function: gender, age, unobserved productivity type, distance to nearest city.

Grade failure probability function: age, gender, grade level, unobserved type

S-P Implementation

Child wage offer function: gender, age, unobserved productivity type, distance to nearest city.

Grade failure probability function: age, gender, grade level, unobserved type

The model is estimated only on control households (landless, nuclear) pre- and post-program and on treatment households prior to the program.

S-P Implementation

Child wage offer function: gender, age, unobserved productivity type, distance to nearest city.

Grade failure probability function: age, gender, grade level, unobserved type

The model is estimated only on control households (landless, nuclear) pre- and post-program and on treatment households prior to the program.

Estimation is by simulated maximum likelihood.

Model Predictions - PROGRESA

Predicted Effect of PROGRESA on School Attendance Rates at Ages 12-15:
NS-NP and S-P Estimation Based on Control Households

	Girls	Boys
Original Subsidy		
NP-NS[1]		
Single child model	.060	.056
	(90%)[2]	(86%)

1. Todd and Wolpin (2010) – two dimensional kernel regression estimator using a bi-weight kernel function.
2. Percent in overlap regions in parentheses.

Model Predictions - PROGRESA

Predicted Effect of PROGRESA on School Attendance Rates at Ages 12-15:
NS-NP and S-P Estimation Based on Control Households

	Girls	Boys
Original Subsidy		
NP-NS[1]		
Single child model	.060	.056
	(90%)[2]	(86%)
Multiple children model	.070	.059
	(68%)	(64%)

1. Todd and Wolpin (2010) – two dimensional kernel regression estimator using a bi-weight kernel function.
2. Percent in overlap regions in parentheses.

Model Predictions - PROGRESA

Predicted Effect of PROGRESA on School Attendance Rates at Ages 12-15:
NS-NP and S-P Estimation Based on Control Households

	Girls	Boys
Original Subsidy		
NS-NP[1]		
Single child model	.060	.056
	(90%)[3]	(86%)
Multiple children model	.070	.059
	(68%)	(64%)
S-P[2]	.064	.077

1. Todd and Wolpin (2010) – two dimensional kernel regression estimator using a bi-weight kernel function.
2. Todd and Wolpin (2006) – DCDP model estimated by simulated maximum likelihood.
3. Percent in overlap regions in parentheses.

A limitation of large scale social experiments, such as PROGRESA, is that it is costly to vary the experimental treatments as a way of evaluating other policies of interest.

Doubling the PROGRESA Subsidy

The Predicted Effect of Alternative Subsidy Policies on School Attendance Rates: NS-NP and S-P Estimation

	Girls	Boys
Original Subsidy		
NP-NS[1]		
Single child	.060	.056
	(90%)	(86%)
2 x Original Subsidy		
NP-NS[1]		
Single child	.141	.116
	(50%)	(53%)

1. Percent in overlap regions in parentheses

Doubling the PROGRESA Subsidy

The Predicted Effect of Alternative Subsidy Policies on School Attendance Rates: NS-NP and S-P Estimation

	Girls	Boys
Original Subsidy		
NP-NS[1]		
Single child	.060	.056
	(90%)	(86%)
Multiple children	.070	.059
	(68%)	(64%)
2 x Original Subsidy		
NP-NS[1]		
Single child	.141	.116
	(50%)	(53%)
Multiple Children	.089	.078
	(50%)	(47%)

1. Percent in overlap regions in parentheses

Doubling the PROGRESA Subsidy

The Predicted Effect of Alternative Subsidy Policies on School Attendance Rates: NS-NP and S-P Estimation

	Girls	Boys
Original Subsidy		
NS-NP[1]		
Single child	.060	.056
	(90%)	(86%)
Multiple children	.070	.059
	(68%)	(64%)
S-P	.064	.077
2 x Original Subsidy		
NS-NP[1]		
Single child	.141	.116
	(50%)	(53%)
Multiple Children	.089	.078
	(50%)	(47%)
S-P	.146	.159

1. Percent in overlap regions in parentheses

The structural-parametric approach permits a quantitative *ex ante* evaluation of a variety of additional policies.

The Effectiveness and Cost of Alternative Education Policies:
Structural-Parametric Estimation

	Baseline	Compulsory Attendance	PROGRESA Subsidy
Mean Completed Schooling			
Girls	6.29	8.37	6.83
Boys	6.42	8.29	6.96
Cost Per Family	0	?	26,096

The Effectiveness and Cost of Alternative Education Policies:
Structural-Parametric Estimation

	Baseline	Compulsory Attendance	PROGRESA Subsidy	Modified Subsidy
Mean Completed Schooling				
Girls	6.29	8.37	6.83	6.97
Boys	6.42	8.29	6.96	7.07
Cost Per Family	0	?	26,096	25,193

The Effectiveness and Cost of Alternative Education Policies:
Structural-Parametric Estimation

	Baseline	Compulsory Attendance	PROGRESA Subsidy	Modified Subsidy	Bonus for Completing 9th Grade
Mean Completed Schooling					
Girls	6.29	8.37	6.83	6.97	6.50
Boys	6.42	8.29	6.96	7.07	6.58
Cost Per Family	0	?	26,096	25,193	36,976

The Effectiveness and Cost of Alternative Education Policies:
Structural-Parametric Estimation

	Baseline	Compulsory Attendance	PROGRESA Subsidy	Modified Subsidy	Bonus for Completing 9th Grade	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
Cost Per Family	0	?	26,096	25,193	36,976	?

The Effectiveness and Cost of Alternative Education Policies:
Structural-Parametric Estimation

	Baseline	Compulsory Attendance	PROGRESA Subsidy	Modified Subsidy	Bonus for Completing 9th Grade	Build Schools	Subsidy + 25% Increase In Child Wage
Mean Completed Schooling							
Girls	6.29	8.37	6.83	6.97	6.50	6.39	6.75
Boys	6.42	8.29	6.96	7.07	6.58	6.55	6.79
Cost Per Family	0	?	26,096	25,193	36,976	?	25,250

Model Validation

1. Tests of within-sample model fit

   NS-NP: not formally valid in light of model pre-testing
   (pre-testing of covariates used for matching)

   S-P: not formally valid in light of model pre-testing
   (structural data-mining)

Model Validation

1. Tests of within-sample model fit
   - NS-NP: not formally valid in light of model pre-testing (pre-testing of covariates used for matching)
   - S-P: not formally valid in light of model pre-testing (structural data-mining)

2. Robustness of findings to assumptions
   - NS-NP: limited set of model alternatives – lots of possible covariates
   - S-P: too many assumptions to be a feasible approach

Model Validation

3. External Validation

A. Regime Shift – McFadden (1977)

- Estimation conducted on sample in one regime - old policy

- Validation conducted on sample in another regime - new policy not available for estimation.

Model Validation

B. Randomized Social Experiment - Wise (1985), Lise, Seitz and Smith (2006), Todd and Wolpin (2006, 2010), Duflo, Hanna and Ryan (2009)

- Estimation conducted on the randomly selected control (treatment) group

- Validation conducted on the randomly selected treatment (control) group – holdout sample

Model Validation

C. Non-Random Holdout Samples - Lumsdaine, Stock and Wise (1992), Keane and Wolpin (2006), French and Jones (2007)

- Estimation conducted on part of the sample, non-randomly chosen – the “control” group

- Validation conducted on the rest of the sample – the “treatment” group.

External Validation - PROGRESA

External Validation of the NS-NP and S-P Estimates
A Comparison of Actual and Predicted School Attendance Rates

	Girls		Boys	
	Actual	Predicted	Actual	Predicted
NS – NP[1]	.090	.070	.033	.059
	(.023)	(.031)[2]	(.022)	(.036)

1. Todd and Wolpin (2010)
2. Bootstrap standard errors

External Validation - PROGRESA

External Validation of the NS-NP and S-P Estimates
A Comparison of Actual and Predicted School Attendance Rates

	Girls		Boys	
	Actual	Predicted	Actual	Predicted
NS – NP[1]	.090	.070	.033	.059
	(.023)	(.031)[2]	(.022)	(.036)
S-P[3]	.080	.064	.038	.077
	(.046)		(.042)	

1. Todd and Wolpin (2010)
2. Bootstrap standard errors
3. Todd and Wolpin (2006)

Model Selection

An open question (work in progress):

Model Selection

An open question (work in progress):

Is a Holdout Sample useful for Model Selection?

Model Selection

An open question (work in progress):

Is a Holdout Sample useful for Model Selection?

To provide context, consider the policy-maker who has designed the PROGRESA experiment – that is, chosen the subsidy schedule, the control and treatment groups, etc.

Model Selection

An open question (work in progress):

Is a Holdout Sample useful for Model Selection?

To provide context, consider the policy-maker who has designed the PROGRESA experiment – that is, chosen the subsidy schedule, the control and treatment groups, etc.

The policy maker would like to know the impact of other subsidy schedules – for example, doubling the subsidy.

Model Selection

The policy-maker would like to obtain the “best” estimate of the effect of the proposed new policy – doubling the subsidy.

Model Selection

The policy-maker would like to obtain the “best” estimate of the effect of the proposed new policy – doubling the subsidy.

The instrument available to the policy maker is the type of data to give to researchers to develop models that can provide an estimate of the policy effect.

Model Selection

The policy-maker would like to obtain the “best” estimate of the effect of the proposed new policy – doubling the subsidy.

The instrument available to the policy maker is the type of data to give to researchers to develop models that can provide an estimate of the policy effect.

1. Data on both the control and treatment households.

Model Selection

The policy-maker would like to obtain the “best” estimate of the effect of the proposed new policy – doubling the subsidy.

The instrument available to the policy maker is the type of data to give to researchers to develop models that can provide an estimate of the policy effect.

1. Data on both the control and treatment households.

2. Data on only one group.

Model Selection

The policy-maker would like to obtain the “best” estimate of the effect of the proposed new policy – doubling the subsidy.

The instrument available to the policy maker is the type of data to give to researchers to develop models that can provide an estimate of the policy effect.

1. Data on both the control and treatment households.

2. Data on only one group.

3. A random sample from each group.

Model Selection

Why not give researchers all of the data (both control and treatment data) and then model average?

Model Selection

Why not give researchers all of the data (both control and treatment data) and then model average?

Let $\theta = \overline{s}_{2\tau} - \overline{s}_0$ be the effect of a doubling of the subsidy on the school attendance rate.

Model Selection

Why not give researchers all of the data (both control and treatment data) and then model average?

Let $\theta = \overline{s}_{2\tau} - \overline{s}_0$ be the effect of a doubling of the subsidy on the school attendance rate.

Let $\widehat{\theta}_{m,d}$ be the estimated subsidy effect based on a model m, for example, $m \in \{NS\text{-}NP, S\text{-}P\}$, given data d.

Model Selection

Then calculate,

$$\widehat{\theta}_d = \sum_m \widehat{\theta}_{m,d} \pi(m|d)$$

and

$$\pi(m|d) = \frac{\pi(d|m)\pi(m)}{\pi(d)}$$

where $\pi(d|m)$ is the marginal likelihood function for model m and $\pi(m)$ is the prior attached by the policy maker to model m.

Model Selection

Researchers want their model to be given the greatest weight in the model averaging. Researchers will data mine.

Model Selection

Researchers want their model to be given the greatest weight in the model averaging. Researchers will data mine.

The policy maker must take that into account, that is, must discount the value of $\pi(d|m)$ given by the researcher.

Model Selection

The trade-off:

Holding out part of the sample from researchers and requiring them to provide a forecast of the hold-out sample data reduces the incentive to data mine as the researcher must be cognizant of overfitting the sample data.

Model Selection

The trade-off:

Holding out part of the sample from researchers and requiring them to provide a forecast of the hold-out sample data reduces the incentive to data mine as the researcher must be cognizant of overfitting the sample data.

However, holding out part of the sample reduces the precision of the policy estimates.

Model Selection

The trade-off:

Holding out part of the sample from researchers and requiring them to provide a forecast of the hold-out sample data reduces the incentive to data mine as the researcher must be cognizant of overfitting the sample data.

However, holding out part of the sample reduces the precision of the policy estimates.

This can be formalized.

Concluding Remarks

There has been significant progress in the development of *ex post* policy evaluation methodologies.

There has been much less attention paid to the development of methodologies for performing *ex ante* policy evaluation. The payoff to further research should be large.