Structural Estimation and Policy Evaluation in Development

Petra E. Todd[1]

[1]University of Pennsylvania

May, 2018

Overview

- Ex post vs. ex ante approaches to policy evaluation
- Use of behavioral models for ex-ante evaluation
- Parametric assumptions not necessarily required
- Static vs. dynamic frameworks
- Applications
- Opportunities for model validation

Ex Post Evaluation

- Goal is to evaluate impacts of an existing program
- Data on a treated group and on a comparison group
- Alternative approaches
    - Randomization
    - Difference-in-difference
    - Matching
    - Regression-discontinuity
    - Control function methods
    - IV methods, MTE, LATE
    - Estimation of a behavioral model

Ex Ante Evaluation

- Evaluate effects of changing parameters of an existing program
- Evaluate the impact of a new program prior to its implementation
  - Needed for optimal program design and placement, which requires simulating program effects and costs
- Evaluate effects of longer terms of exposure to an existing program than are observed in the data

Some Examples of Ex Ante Evaluations Using Static Models

- Forecast demand for a new good introduced into the choice set
  - e.g. McFadden (1977) - BART subway
- Forecast effect of changing the characteristics of a good on consumer demand
  - Berry,Levinsohn and Pakes (1995) - changing car characteristics (e.g. price, fuel efficiency)
  - BLP model often used to analyze effects of mergers

Some Examples of Ex Ante Evaluations Using Dynamic Models

- Wise (1985): Predict the effect of housing subsidy on housing demand
- Lumsdaine, Stock and Wise (1992): Predict the effect of retirement bonus on retirement patterns
- Lise, Seitz and Smith (2003) - Predict effects of welfare bonus program on job search
- Todd and Wolpin (2006) - Predict effects of school subsidy program on school attendance and work behaviors

The Importance of Economic Models in Ex Ante Policy Evaluation

- Koopmans (1947), Marschak (1953), Hurwicz (1962)
  - Recognize that an economic model provides a way of extrapolating from historical experience
  - Observe that it is not necessary to know the entire structure of the problem to answer certain policy questions (e.g. tax changes)

Recent Efforts at Nonparametric Ex Ante Evaluation

- Ichimura and Taber (1998, 2002)
  - Present general set of conditions under which nonparametric policy evaluation is possible
  - Estimate effects of a tuition subsidy using tuition variation in the data
- Heckman (2000,2001)
  - Discusses "Marshak's Maxim," provides new examples of where nonparametric assessment of new policies is feasible
- Blomquist and Newey (2002)
  - Nonparametric estimation of labor supply responses with nonlinear budget sets.
- Todd and Wolpin (2010)
  - Nonparametric estimation of effects of school and income subsidies on school attendance

Evaluate Effects of School Attendance Subsidy When Child Wage Offers are Observed (Todd and Wolpin (2010))

- Household makes a single period decision about whether to send a child to school or work
- Utility depends on consumption (c) and on whether the child attends school (indicator s).
- A child not attending school works at wage w.
- y denotes household income, net of the child's earnings
- The household solves the problem:

$$
\begin{aligned}
& \max_{\{s\}} \; U(c, s, \mu) \\
& s.t. \\
c \;=\; & y + w(1 - s).
\end{aligned}
$$

The optimal choice $s^* = \varphi(y, w, \mu)$, where μ denotes unobservable preference heterogeneity.

- Consider a policy that provides a subsidy τ for school attendance. The problem becomes:

$$\begin{aligned} & \max_{(s)} \; U(c, s, \mu) \\ & s.t. \\ c \;\; = \;\; & y + w(1-s) + \tau s. \end{aligned}$$

The constraint can be rewritten as

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

which shows that the optimal choice of s in the presence of the subsidy is $s^{**} = \varphi(\tilde{y}, \tilde{w}, \mu)$, where $\tilde{y} = y + \tau$ and $\tilde{w} = w - \tau$.

Estimation

- Under the assumption that :

$$f(\mu|y,w) = f(\mu|\tilde{y},\tilde{w}),$$

  Can estimate the effect of the subsidy program on the proportion of children attending school by comparing children from families with income $\tilde{y}$ and child wage offers $\tilde{w}$ to children from families with income y and child wages w.
- Clearly a stringent condition.
- To make more plausible, could condition on a vector of family characteristics, x, and assume:

$$f(\mu|y,w,x) = f(\mu|\tilde{y},\tilde{w},x).$$

Estimation

- A matching estimator of average program effects for those offered the program (the "intent-to-treat" estimator):

$$\frac{1}{n}\sum_{\substack{j=1 \\ j,i\in S_P}}^{n}\{E(s_i|w_i=w_j-\tau,y_i=y_j+\tau)-s_j(w_j,y_j)\},$$

  where $s_j(w_j,A_j)$ denotes the school attendance decision for a child of family j with characteristics (w_j,y_j).
- The average can only be taken over the region of overlapping support S_P, which in this case is over the set of families j for which the values $w_j-\tau$ and $y_j+\tau$ lie within the observed support of w_i and y_i.
- $E(s_i|w_i=w_j-\tau,y_i=y_j+\tau)$ can be estimated by nonparametric regression.

- We can evaluate the effects of a range of school subsidy programs.
- Nonparametric ex ante policy evaluation is feasible even when there is no variation in the data in the policy instrument (here, the price of schooling).

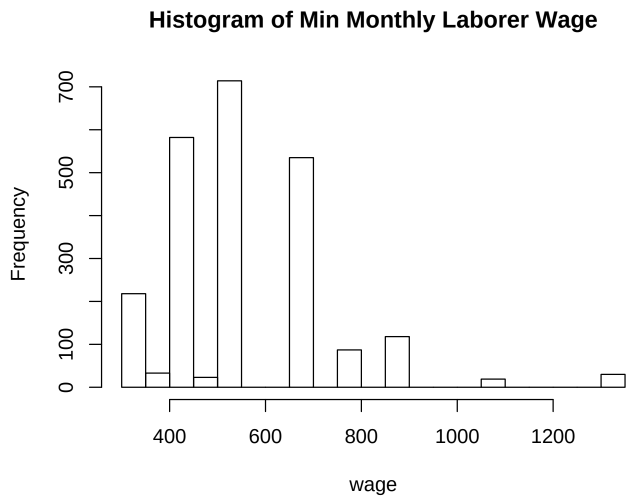
Application: The PROGRESA Program

- Large scale anti-poverty program
  - Begun in 1997, now has budget of about 1 billion US Dollars
  - About 20% of Mexican families participating
- Provides educational grants to mothers to encourage children's school attendance
- Benefit levels increase with grade level, higher for girls
- Subsidies amount to about 20% of average annual income
- Data from the initial rural evaluation that randomized 506 villages in or out of the program.

Table 1
Monthly Transfers for School Attendance

School Level	Grade	Gender	
		Female	Male
Primary	3	70	70
	4	80	80
	5	105	105
	6	135	135
Secondary	7	210	200
	8	235	210
	9	235	225

Figure 1


Histogram of Min Monthly Laborer Wage
Frequency
0
100
300
500
700
400
600
800
1000
1200
wage

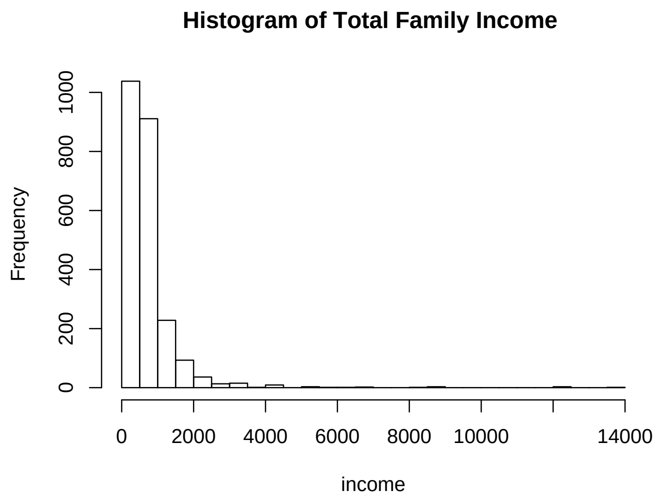

Histogram of Total Family Income
Frequency
0
200
400
600
800
1000
0
2000
4000
6000
8000
10000
14000
income

Table 2(a)
Comparison of Ex-Ante Predictions and Experimental Impacts
Multiple-child model (Bootstrap standard errors in parentheses) †

	Boys			
Ages	Experimental	Predicted	Sample-Sizes‡	% overlapping support
12-13	0.05** (0.02)	0.05 (0.03)	374, 610	68%
14-15	0.02 (0.03)	0.09* (0.05)	309, 569	61%
12-15	0.03 (0.02)	0.06** (0.03)	683, 1179	64%
	Girls			
	Experimental	Predicted	Sample-Sizes‡	% overlapping support
12-13	0.07 (0.07)	0.04 (0.04)	361, 589	67%
14-15	0.11** (0.04)	0.11* (0.06)	361, 591	68%
12-15	0.09 ** (0.02)	0.07** (0.04)	677, 1180	68%
	Boys and Girls			
	Experimental	Predicted	Sample-Sizes‡	% overlapping support
12-13	0.06** (0.02)	0.04 (0.03)	735, 1199	67%
14-15	0.07** (0.03)	0.10** (0.04)	625, 1160	64%
12-15	0.06** (0.02)	0.07** (0.02)	1360, 2359	66%

†Standard errors based on 500 bootstrap replications. Bandwidth equals 200 pesos. Trimming implemented using the 2% quantile of positive density values as the cut-off point.
‡The first number refers to the total control sample and the second to the subset of controls that satisfy the PROGRESA eligibility criteria.

Table 2(b)
Effects of Counterfactual Subsidy Levels
Multiple-child model (% in overlap region in parentheses)

	Boys		
Ages	2* Original	Original	0.75*Original
12-13	0.01 (50%)	0.05 (68%)	0.01 (92%)
14-15	0.16 (43%)	0.09 (61%)	0.04 (93%)
12-15	0.08 (47%)	0.06 (64%)	0.02 (93%)
	Girls		
	2* Original	Original	0.75*Original
12-13	0.04 (48%)	0.04 (67%)	0.04 (93%)
14-15	0.15 (52%)	0.11 (68%)	0.04 (93%)
12-15	0.09 (50%)	0.07 (68%)	0.04 (93%)
	Boys and Girls		
	2* Original	Original	0.75*Original
12-13	0.03 (49%)	0.04 (67%)	0.02 (93%)
14-15	0.15 (48%)	0.10** (64%)	0.04 (93%)
12-15	0.08 (49%)	0.07** (66%)	0.03 (93%)

† Bandwidth equals 200 pesos. Trimming implemented using the 2% quantile of positive density values as the cut-off point.

Table 4
Predicted Impact of an Unconditional Income Transfer in the Amount of 5000 pesos/year
Multiple-child model (Bootstrap standard errors in parentheses) †

	Boys		
Ages	Predicted	Sample-Sizes‡	% overlapping support
12-13	-0.02 (0.03)	374, 610	89%
14-15	-0.06 (0.05)	309, 569	90%
12-15	-0.04 (0.03)	683, 1179	89%
	Girls		
	Predicted	Sample-Sizes‡	% overlapping support
12-13	-0.03 (0.04)	361, 589	88%
14-15	0.00 (0.05)	316, 591	88%
12-15	-0.02 (0.03)	677, 1180	88%
	Boys and Girls		
	Predicted	Sample-Sizes‡	% overlapping support
12-13	-0.03 (0.03)	735, 1199	88%
14-15	-0.03 (0.03)	625, 1160	89%
12-15	-0.03 (0.02)	1360, 2359	89%

†Standard errors based on 500 bootstrap replications. Bandwidth equals 200 pesos. Trimming implemented using the 2% quantile of positive density values as the cut-off point.
‡The first number refers to the total control sample and the second to the subset of controls that satisfy the PROGRESA eligibility criteria.

Limitations of Nonparametric Policy Evaluation

- Not possible when there is an alternative use of children's time, such as leisure.
- Imposes strong assumptions on the distribution of unobservables.
- Child wage offers usually not observed when children are not working.

Parametric Static Model with Unobserved Heterogeneity and Partial Wage Observability

- A couple chooses between sending their child to work ($d_{it} = 1$) or school ($d_{it} = 0$)
- Utility is

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

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:

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

- x_{it} ($\subseteq X_{it}$) include, perhaps, parents' schooling or the child's gender.
- ε_{it} is an iid random preference shock to the utility of the child's school attendance (iid assumption can be relaxed)

- The child receives a wage offer of w_{it} and the household otherwise generates income y_{it}.
- The budget constraint is

$$C_{it} = y_{it} + w_{it} d_{it} ,$$

where there are assumed to be no costs associated with attending school.

- Wage offers only observed for children who work (partial observability), so we also need a wage offer equation:

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

- z_{it} ($\subseteq Z_{it}$) would contain, for example, the child's age, gender, or factors affecting the demand for child labor, such as distance to a city.
- η_{it} is an iid wage shock
- We do not include the child's current educational attainment in z to maintain the static nature of the model.

- Alternative-specific utilities, U_{it}^1 if the child works and U_{it}^0 if the child attends school as

$$
\begin{aligned}
U_{it}^1 &= y_{it} + w_{it} \;, \\
U_{it}^0 &= y_{it} + x_{it}\beta + \varepsilon_{it}.
\end{aligned}
$$

- Substituting the wage equation yields U_{it}^1 - U_{it}^0

$$
\begin{aligned}
v_{it}^*(x_{it}, z_{it}, \varepsilon_{it}, \eta_{it}) &= z_{it}\gamma - x_{it}\beta + \eta_{it} - \varepsilon_{it} \\
&= \xi_{it}^*(\Omega_{it}^-) + \xi_{it},
\end{aligned}
$$

where $\xi_{it} = \eta_{it} - \varepsilon_{it}$, $\xi_{it}^*(\Omega_{it}^-) = z_{it}\gamma - x_{it}\beta$ and Ω_{it}^- consists of z_{it} and x_{it}.

Estimation: Likelihood Function

- The likelihood function, incorporating the wage information, is

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

Ex Ante Evaluation: Predict Effects of a Subsidy

- Assume that f(ε, η) is joint normal with variance-covariance matrix, $\Lambda = \begin{pmatrix} \sigma_\varepsilon^2 & \cdot \\ \sigma_{\varepsilon\eta} & \sigma_\eta^2 \end{pmatrix}$.
- Parameters to be estimated include β, γ, π, σ_ε^2, σ_η^2, and $\sigma_{\varepsilon\eta}$.
- Joint normality is sufficient to identify the wage parameters (γ and σ_η^2) as well as (σ_η^2 - $\sigma_{\varepsilon\eta)}$)/$\sigma_\xi$ (Heckman 1979).

- The probability that the child works is

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

  where Φ is the standard normal cumulative distribution.
- Data on work choices identify γ / σ_ξ and β / σ_ξ.
- To identify σ_ξ, there are three types of variables: - variables only in z (in the wage function), - variables only in x (in the utility function), and - variables in both x and z.
- Having identified the $\gamma' s$, the identification of σ_ξ (and thus also $\sigma_{\varepsilon\eta}$) requires at least one variable only in the wage equation.
- For example, a variable that affects the demand for labor but does not affect the utility value the couple places on the child's school attendance.

Predict effects of a subsidy

- Suppose the government wants to predict the effects of a schooling subsidy
- With the subsidy τ

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

- It is necessary to have identified σ_ξ to predict the effects of the subsidy
- Government outlays on the program equal the number of children times the probability of attending school.
- Can study effects of a range of subsidies.
- Exogenous variation in the wage (independent of utility) is crucial for identification.

Ex Ante Evaluation Using Dynamic Models

- In the static model, there was no connection between the current period decision and future utility.
- Suppose that child's wage increases with work experience

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

where $h_{it} = \sum_{\tau=1}^{\tau=t-1} d_{i\tau}$ is work experience at the start of period t.

- Alternatively, parents' utility could depend on the number of school years completed, so that current attendance affects future utility.

Dynamic Model continued

- The couple maximizes the PDV of remaining lifetime utility starting from t=1 and ending at T.
- $V_t(\Omega_{it})$ denotes the maximum expected present discounted value of remaining lifetime utility at t given the state space and discount factor δ,
- 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.

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

- With the wage equation, h_{it} becomes part of the state space and evolves according to $h_{it} = h_{i,t-1} + d_{i,t-1}$

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

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

each of which obeys the Bellman equation

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

- The expectation is taken over the distribution of the random components of the state space at t+1 conditional on the state space elements (here the shocks are mutually serially independent.)

- The latent variable in the dynamic case is $V_t^1(\Omega_{it}) - V_t^0(\Omega_{it})$:

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

- A full solution of the dynamic programming problem consists of finding E[max($\mathrm{V}_t^0(\Omega_{it})$, $\mathrm{V}_t^1(\Omega_{it})$)] at all values of Ω_{it}^-, denoted by Emax(Ω_{it}^-), for all t=1,...,T.
- Same as static case, except now includes the difference in the future component of the expected value functions under the two alternatives.

Estimation: Likelihood function

- Assume researcher has data from t_{1i} to t_{Li}.

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

- 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$_{i\tau}\gamma_1$ - γ_2 h$_{it}$) and Pr(d$_{i\tau}$ = 0) = 1 - Pr($\xi_{i\tau} \geq -\xi_{i\tau}^{*}(\Omega_{i\tau}^{-})$).
- If the error is not additive, then calculating the joint regions of the error that determine the probabilities that enter the likelihood can be done numerically.

Extension to Multinomial Choice

- If there are K>2 mutually exclusive alternatives, there will be K-1 latent variable functions (relative to one of the alternatives, arbitrarily chosen).
- 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 at all values of Ω_{it}^- and at all t, is computationally more intensive.
- Defining d_{it}^n as the discrete {0,1} choice variable corresponding to the nth choice ($n = 1,..,N$) and $\widetilde{d}_{it}$ as the N element vector of those choices, there would be at most $K = 2^N$ mutually exclusive choices.

Allowing for Permanent Unobserved Heterogeneity

- In the example, unobservables were iid, but serial dependence is feasible.
- A standard specification assumes that agents can be distinguished, in terms of preferences and opportunities, by a fixed number of types. (Similar to approach of Heckman and Singer, 1981, in duration analysis)
- 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 $w_{ijt} = \gamma_{oj} + z_{it}\gamma_1 + \gamma_2 h_{it} + \eta_{it}$.
- The dynamic program must then be solved for each type and the likelihood function is a weighted average over each type in the sample.
- Type proportions are estimated along with the other parameters.

Applications: Education

- Todd and Wolpin (2006), Attanasio, Meghir and Santiago (2005) - analyze effects of CCT programs in *Mexico* on school-going, child labor and fertility.
- Baird, Ozler, Shapira and Todd (2017) - analyze effects of CCT and UCT programs in *Malawi* on schooling-going, marriage and fertility of adolescent girls
- Duflo, Hanna and Ryan (2008) - analyze effects of a program in *India* that provides financial incentives to teachers to not be absent from school
- Bravo, Mukhopadyay and Todd (2009) - examine how the introduction of the school voucher program in *Chile* affects life-cycle education and working decisions
- Behrman, Tincani, Todd and Wolpin (2016) - Explore how teacher wage policies in *Chile* affect the decision to become a teacher and teacher's decisions whether to work in the public, private-voucher, or private non-voucher sectors.

Applications: Pension systems

- McKee (2006) - Analyzes labor supply for older men in *Indonesia* incorporating own labor, family transfers and coresidence. Uses the model to study effects of introducing two new types of public pensions.
- Velez-Grajales (2009), Joubert (2010), Joubert and Todd (2015) - Analyze effects of reforms to the pension program in *Chile* on household decision-making with regard to savings, labor supply (in formal and informal sectors) and retirement.

Applications: Business savings/investment

- *India*: Rosenzweig and Wolpin (1993) - estimates a model of agricultural investment behavior that incorporates income uncertainty, borrowing constraints, and investment in assets (bullocks). The model is used to analyze welfare effects of policies that provide regular income and that insure farmers against weather shocks.
- Kaboski and Townsend (2007) - analyzes the effect of *Thailand's* "Million Baht Village Fund" program on short-term credit, consumption, agricultural investment, and income growth (from business and labor) and on asset growth. Compares this program to a hypothetical CCT program.
- Wang (2015) - analyzes the impact of the expansion of formal credit on household credit market choices, where households can also borrow and lend in an informal credit market.

Applications: Migration

- Colussi (2006) - analyzes effect of migration laws and immigration enforcement levels on migration from rural *Mexico* to the US.
- Kirdar (2009) - examines the determinants of return migration from Germany for immigrants from *Greece, Italy, Spain and Turkey*. Examines whether decisions are driven by reaching savings accumulation goals.
- Adda, Dustmann and Gorlach (2014) - studies the immigration and return migration decisions of Turkish immigrants in Germany. Estimates a model describing labor market participation, wages, return decisions and accumulation of two forms of human capital, work experience and cultural integration.
- Lagakos, Mobarak, Waugh (2017) - studies welfare effects of using subsidies to encourage rural-urban migration in *Bangladesh*.

Applications: Fertility/Marriage

- Wolpin (1984) - analyzes life-cycle fertility in an environment where infant survival is uncertain. Analyzes effects on the timing, spacing and number of children using data from *Malaysia*.
- Shapira (2013) - Analyzes how HIV testing programs in *Malawi* that change women's subjective belief's about their HIV status affect fertility behavior.
- Luis Garcia (2017) - Analyzes effects of the one-child policy in *China* on life-cycle fertility behavior. The policy imposed fines for having more than one child. The level of the fines varied geographically, over time and with the family's income.

Applications: The Production Function for Child Development

- Azuero (2015) - Estimates a collective model of parents' labor supply and investment in skills for children. The model incorporates a dynamic skill formation production function. It is used to simulate the effects of alternative cash and in-kind transfer policies designed to increase skill investment.