Empirical Methods for Policy Evaluation II

Matteo Bobba

Toulouse School of Economics (TSE)

TSE PhD Program
Fall 2021

Part 2: Dynamic Discrete Choice Models & RCTs

- **Context: implemented policy and evaluation**
- Methods
  - Keane, Todd and Wolpin (HLE, 2011)
- Applications
  - Todd and Wolpin (AER, 2006)
  - Attanasio, Meghir and Santiago (ReStud, 2012)

The *Progresa* Program in Mexico

- Large scale anti-poverty program.
  - Began in 1997 in rural areas and rapidly expanded throughout the country
  - About 20% of Mexican families participating
- Provides educational grants to mothers to encourage children's school attendance (among other things...)
  - Benefits levels increase with grades attained, higher for girls
  - Subsidies amount to about 20% of average annual income
- Data from the initial rural evaluation of the program
  - Randomized phase-in design at the village level
  - 506 villages, 320 treated (early start, in March 1998) and 186 control (late start, November 1999)
  - Within villages, both eligible and non-eligible HHs (wealth index)

Progresa: Experimental Evidence

- The program had positive impacts on:
  - Nutrition and growth
  - Consumption and poverty
  - Secondary school enrollment (no impact on primary and no effect on learning outcomes)
- Spillover effects
  - From program beneficiaries to non-beneficiaries
  - Among program beneficiaries
- What are the mechanisms? Lets take the schooling decision
  - Changes in the relative price of schooling+income effect
  - Peer effects
  - GE effect on child wages

Part 2: Dynamic Discrete Choice Models & RCTs

- Context: implemented policy and evaluation
- Methods
  - **Keane, Todd and Wolpin (HLE, 2011)**
- Applications
  - Todd and Wolpin (AER, 2006)
  - Attanasio, Meghir and Santiago (ReStud, 2012)

Why a Dynamic Model to Evaluate a Cash Transfer Program?

- In the static model, there is no connection between the current period decision and future utility
- Child's wage may increase with past work experience
- Past education could change attitudes towards attendance
- Parents' utility may depend on the number of school years completed, so that current attendance affects future utility
- The grant itself creates dynamics because not going to school one year reduces the total number of years the child can be subsidized: the grant is only available until 17

Dynamic Discrete Choice (DDC): General Framework

- At each discrete period t, an individual chooses among K mutually exclusive alternatives

$$d_k(t) = \begin{cases} 1 & \text{if } k \text{ is chosen} \\ 0 & \text{otherwise} \end{cases}$$

- Individual utility in period t depends on the vector of state variables $s_t = (x_t, \epsilon_t)$. Baseline assumptions:
  1. $U(d_t, x_t, \epsilon_t) = U(d_t, x_t) + \epsilon_t(d)$, where $\epsilon_t(d) = \sum_{k \in K} d_k(t)\epsilon_k(t)$
  2. ϵ_t are iid across agents and over time $\sim F_\epsilon(\epsilon_t)$
  3. $F_x(x_{t+1}|d_t, x_t, \epsilon_t) = F_x(x_{t+1}|d_t, x_t)$

Dynamic Discrete Choice (DDC): Value Functions and Bellman Principle

- In each t individuals' optimal choices satisfy

$$V(s,t) = \max_{(d_t, d_{t+1}, ..d_T)} \mathbb{E}\left[\sum_{\tau=t}^{T} \beta^{\tau-t} \sum_{k=1}^{K} d_k(\tau) U(d_\tau, s_\tau) | s_\tau\right]$$

- Or, alternatively $V(s,t) = \max\{V_1(s_t,t), ..., V_K(s_t,t)\}$
- Where $V_k(s,t) = U(d_t, s_t) + \beta \mathbb{E}\left[V(s_{t+1}, t+1) | s_t, d_k(t) = 1\right]$ and $V_k(s,T) = U(d_T, s_T)$

Dynamic Discrete Choice (DDC): The Curse of Dimensionality

- The solution of the dynamic programming problem requires to compute $V_k(s,t)\ \forall(k,s,t)$
  - Estimation requires that the dynamic programming problem be solved many times
  - The computational cost increases exponentially with the number of state variables
- Three approaches
  1. Exact full-solution methods (Rust, 1987)
  2. Non-full solution methods (Hotz and Miller, 1993)
  3. Approximation/interpolation (Keane and Wolpin, 1994)

Part 2: Dynamic Discrete Choice Models & RCTs

- Context: implemented policy and evaluation
- Methods
  - Keane, Todd and Wolpin (HLE, 2011)
- Applications
  - **Todd and Wolpin (AER, 2006)**
  - Attanasio, Meghir and Santiago (ReStud, 2012)

Todd and Wolpin (AER, 2006)

- Dynamic discrete choice model of children's time allocation and family fertility
- Each year a married couple decides on whether
  1. Each of their children between the ages of 6 and 15 attend school
  2. Each of their children between the ages of 6 and 15 remain at home
  3. Each of their children between the ages of 12 and 15 work in the labor market
  4. The wife becomes pregnant
- Family income is the sum of parental and children earnings, both of which are subject to idiosyncratic time-varying shocks
  - No parental labor supply decisions
  - No saving or borrowing

Model Description

- The household problem is

$$\max_{\{s(t),l(t),p(t)\}} U(t) = U(C(t), p(t), n(t), s_b(t), s_g(t), S_b(t), S_g(t), l_b(t), l_g(t), z_s; \epsilon(t), \mu)$$

$$s.t.\ C(t) = y_p(t) + \sum_n y_o(t, \tau_n) h(t, \tau_n), \text{where:}$$

$$\log y_p(t) = y_p(a_p(t), z_c, \epsilon_{y_p}(t); \mu_{y_p})$$

$$\log y_o(t, \tau_n) = y_o(t - \tau_n, \mathbb{I}(b(\tau_n) = 1), z_c, \epsilon_{y_o}(t); \mu_{y_o})$$

$$\pi_c(t, \tau_n) = \pi(t - \tau_n, S(t, \tau_n) | s(t, \tau_n) = 1, \mu_c)$$

- Unobserved heterogeneity:
  - ϵ-shocks are assumed jointly normal and serially uncorrelated
  - μ-types are known to parents and distributed $g(\mu)$

Dynamic Programming

- Let $d_k(t) = 1$ if the k-the alternative is chosen at t
- Let $\Omega(t)$ be the state space at t
- Value function can be written as:

$$
\begin{aligned}
V(\Omega(t), t) &= \max_{k \in K(t)} \left[V^k(\Omega(t), t) \right], \\
V^k(\Omega(t), t) &= U^k(\Omega(t), t) + \delta \mathbb{E}(V(\Omega(t+1), t+1 | d_k(t) = 1, \Omega(t))) \text{ for } t < \bar{T}, \\
&= U^k(\Omega(\bar{T}), \bar{T}) \text{ for } t = \bar{T}.
\end{aligned}
$$

- Model solved numerically using approximation/interpolation method

Estimation

- Numerical solution of the model delivers the $V^k(t)$ functions
- Let outcome vector at t denoted by $O(t) = \{d^k(t), y_o(t), c(t), y_p(t)\}$
- Likelihood for a sample of N households beginning at marriage $t = t_{mn}$ and ending at some $t = \bar{t}_n$ is

$$\prod_{n=1}^{N}\sum_{j=1}^{J} Pr(O(\bar{t}_n), ..., O(t_{mn})|\bar{\Omega}(t_{mn}), \text{type} = j) Pr(\text{type} = j|\bar{\Omega}(t_{mn}))$$

- $\bar{\Omega}(t_{mn})$ is the initial state space net of family type and stochastic shocks
- Measurement error in children's wages: $y_o^{obs} = y_o(t) exp(\eta(t))$

Aside: The Initial Conditional Problem in Dynamic Discrete Choice Models

- Initial conditions (ages of marriage of both parents and the distances) are assumed exogenous conditional on type μ
- This is fine for families with complete decision histories, while they depend on $O(\bar{t}_n), ..., O(t_{nm})$ through the vector μ for those with incomplete histories
- Approximate $Pr(\text{type} = j|\bar{\Omega}(t_{mn}))$ using a multinomial logit
  - The logit parameters are themselves functions of the structural parameters
  - Identification through functional form assumptions

Estimation Results: Unobserved Types

Figure: Distribution of Observables by Type

	Type 1		Type 2		Type 3	
	Girls	Boys	Girls	Boys	Girls	Boys
Percent of children age 6–11 in school	98.5	99.4	97.6	99.9	78.7	64.2
Percent of children age 12–15 in school	37.3	50.2	84.6	86.9	44.5	36.8
Percent of children age 12–15 at home	55.9	31.0	11.3	7.0	33.5	30.9
Percent of children age 12–15 at work	6.8	18.8	4.1	6.1	21.9	32.3
Mean wage of children 12–15	2,675	3,599	2,600	3,499	2,739	3,666
Mean parental income	9,953		11,944		10,107	
Percent becoming pregnant	15.0		5.6		14.8	
Percent of sample	38.8		52.0		9.2	

Internal Model Validation: Within-Sample Fit

Figure: Actual and Predicted Choice Distribution by Child Age and Gender

Boys							
	Actual			Predicted			
Age	School	Work	Home	School	Work	Home	χ^2
6	0.933	—	0.066	0.923	—	0.077	0.58
7	0.981	—	0.019	0.980	—	0.020	0.02
8	0.987	—	0.013	0.980	—	0.020	0.99
9	0.994	—	0.006	0.979	—	0.021	3.49
10	0.982	—	0.018	0.974	—	0.026	0.86
11	0.977	—	0.023	0.964	—	0.036	1.45
12	0.885	0.021	0.094	0.846	0.039	0.115	3.99
13	0.780	0.084	0.136	0.736	0.078	0.186	4.51
14	0.677	0.157	0.166	0.619	0.191	0.190	3.41
15	0.490	0.276	0.235	0.520	0.251	0.229	0.88
Girls							
6	0.965	—	0.035	0.942	—	0.058	3.84
7	0.976	—	0.024	0.968	—	0.032	0.77
8	0.989	—	0.011	0.976	—	0.024	1.96
9	0.991	—	0.009	0.975	—	0.025	3.26
10	0.979	—	0.021	0.970	—	0.030	0.93
11	0.969	—	0.031	0.948	—	0.052	2.97
12	0.896	0.007	0.097	0.854	0.020	0.126	4.61
13	0.726	0.028	0.245	0.676	0.025	0.299	2.85
14	0.582	0.089	0.329	0.566	0.092	0.342	0.22
15	0.419	0.123	0.458	0.402	0.157	0.442	1.68

Note: χ^2 (0.05, 1) = 3.84, χ^2 (0.05, 2) = 5.99.

Out-of-Sample Model Validation using the Experiment

Figure: Comparison of Ex-Ante Predictions and Experimental Impacts

	Girls age 12–15			Girls age 12–15, behind in school			Girls age 13–15, HGC ≥ 6, behind in school		
	(1) Actual	(2) Pred. with Subsidy	(2)–(1)	(1) Actual	(2) Pred. with Subsidy	(2)–(1)	(1) Actual	(2) Pred. with Subsidy	(2)–(1)
97 Control	65.3	72.7	7.4	58.3	67.0	8.7	40.9	58.6	17.7
98 Control	66.5	72.9	6.4	58.7	66.9	8.2	44.4	60.6	16.2
97 Treatment	62.9	73.0	10.1	56.9	67.6	10.7	30.3	56.2	25.9
Experimental treatment effect:									
Cross section		8.0 (4.6)			12.8 (5.7)			7.1 (8.6)	
Difference-in-difference		10.3 (6.7)			14.1 (8.3)			17.7 (12.0)	

	Boys age 12–15			Boys age 12–15, behind in school			Boys age 13–15, HGC ≥ 6, behind in school		
	(1)	(2)	(2)–(1)	(1)	(2)	(2)–(1)	(1)	(2)	(2)–(1)
97 Control	68.8	79.6	10.8	64.0	75.8	11.8	59.0	72.7	13.7
98 Control	72.5	80.2	7.7	67.4	78.0	10.6	57.1	72.8	15.7
97 Treatment	69.5	79.4	9.9	64.2	75.8	11.6	52.6	71.6	19.0
Experimental treatment effect:									
Cross section		3.8 (4.2)			4.2 (5.2)			1.2 (8.4)	
Difference-in-difference		3.1 (6.1)			4.0 (7.4)			3.8 (11.7)	

Counterfactual Policy Experiments

1. Long-term impact of the subsidy program
    - Fertility outcomes are essentially invariant to the subsidy
    - Small long run effects (compared to short-run) on secondary school attendance rates and completed schooling

2. Alternative subsidy programs
    - Mean completed schooling increases at a linear rate with increments in subsidy amounts up to the original amount and then at a slightly diminishing rate (half-subsidy more cost-effective)
    - Budget neutral change in grade eligibility (only secondary school) increases completed schooling by 25%
    - Rewarding junior secondary school completion instead of attendance decreases primary school completion due to substitution effect between siblings
    - Building a secondary school in each village where it is absent has minor effects on schooling

Part 2: Dynamic Discrete Choice Models & RCTs

- Context: implemented policy and evaluation
- Methods
  - Keane, Todd and Wolpin (HLE, 2011)
- Applications
  - Todd and Wolpin (AER, 2006)
  - **Attanasio, Meghir and Santiago (ReStud, 2012)**

Attanasio, Meghir and Santiago (ReStud, 2012)

- Similar dynamic discrete choice structure with some differences wrt TW
  1. No fertility decision
  2. Binary choice: school vs. work
  3. Each child's utility is maximized independently of that of the parents or of other children
  4. Allow for marginal utility of the subsidy to differ from marginal utility of other sources of income
  5. Allow for equilibrium effects of the program on children's wages

Model Overview

- Utility for child i of attending or not school in time (age) t is:

$$u^s_{it} = \gamma\delta g_{it} + \mu_i + \alpha' z_{it} + \lambda e d_{it} + 1(p_{it} = 1)\beta^p x^p_{it} + 1(s_{it} = 1)\beta^s x^s_{it} + \epsilon_{it}$$
$$u^w_{it} = \delta w_{it}$$

- Unobserved heterogeneity:
    1. ϵ_{it} is an iid logistic shock to costs of schooling
    2. μ_i is drawn from a discrete distribution whose points of supports and probabilities will be estimated empirically
- Child may not be successful in completing a grade (exogenous probability p^s_{it} estimated from the data)

Value Functions

- Terminal value function (returns to schooling):

$$V(ed_{i,18}) = \frac{\alpha_1}{1 + \exp(-\alpha_2 ed_{i,18})}$$

- Value of school and of work at age t can be written as:

$$\begin{aligned} V^s_{it}(ed_{it}|\Omega_{it}) &= u^s_{it} + \beta\{p^s_{it}(ed_{it}+1)\mathbb{E}\max\left[V^s_{it+1}(ed_{it}+1), V^w_{it+1}(ed_{it}+1)\right]\} \\ &\quad + \{(1 - p^s_{it}(ed_{it}+1)\mathbb{E}\max\left[V^s_{it+1}(ed_{it}), V^w_{it+1}(ed_{it})\right]\} \\ V^w_{it}(ed_{it}|\Omega_{it}) &= u^w_{it} + \beta\mathbb{E}\max\left[V^s_{it+1}(ed_{it}), V^w_{it+1}(ed_{it})\right] \end{aligned}$$

- $\mathbb{E}\max$ functions have closed form expressions due to the logistic in $u_{it} = \tilde{u}_{it} + \epsilon_{it}$

Wages and General Equilibrium Responses

- Standard Mincer-type wage equation estimated outside of the schooling model:

$$\begin{aligned} \ln w_{ij} &= q_j + a_1 age_i + a_2 educ_i + a_3 IMR_i + \omega_{ij} \\ q_j &= b_1 \ln w_j^{ag} + b_2 P_j \end{aligned}$$

  - Notice that the wage does not depend on children's education
  - q_j represents the log price of human capital in the locality
  - $\ln w_j^{ag}$ is a sufficient statistic for the overall level of labor demand in the local area (see toy model example in the paper with two inputs, child and adult labor)
  - The program pushes up child wages by decreasing child labor (implied elasticity of about -1.2)
  - No or limited selection on unobserved ability ($\hat{a}_3 < 0$ and very small)

Initial Conditions

- As in TW, data consist of level of education ed_{it} and enrollment decision after the experiment started (i.e. we do not observe the entire history of schooling)
- Past education decisions are correlated with unobserved ability μ_i
- Reduced-form ordered probit model:

$$P(ed_{it} = e \mid z_{it}, x_{it}^s, x_{it}^p, h_i, w_{it}, \mu_i)$$

  - μ_i is added to the normally distributed random variable of the ordered probit, effectively yielding a mixture of normals
  - h_i reflects past schooling costs, such as distance to closest school in the past
  - Past distance to schools should not affect current school participation decisions

Identification and Estimation

- Comparison between treatment and control villages and between eligibles and ineligible HHs within villages identifies the effect of the grant (extensive margin)
- The fact that children of different ages attend the same grade identifies the effect of the size of the grant (intensive margin)
- Parameters of the (joint) model are estimated by maximum likelihood

$$L_i = \int_{\mu} P(at_i = 1 \mid z_{it}, x^s_{it}, x^p_{it}, ed_{it}, w_{it}, \mu_i) P(ed_{it} = e \mid z_{it}, x^s_{it}, x^p_{it}, h_i, w_{it}, \mu_i) dg(\mu_i)$$

- The distribution of unobservables $g(\mu_i)$ is assumed independent of all observables

Estimation Results I

Figure: The distribution of unobserved heterogeneity

	A	B	C
Point of Support 1	−9·706	−8·327	−4·290
	1·041	*1·101*	*2·46*
Point of Support 2	−14·466	−13·287	−17·62
	1·173	*1·208*	*3·144*
Point of Support 3	−5·933	−4·301	−0·267
	0·850	*0·941*	*2·45*
Probability of 1	0·513	0·518	0·490
	0·024	*0·023*	*0·032*
Probability of 2	0·342	0·335	0·270
	0·022	*0·021*	*0·017*
Probability of 3	0·145	0·147	0·240
Load factor for initial condition	0·108	0·102	0·068
	0·016	*0·014*	*0·013*

Notes: Column A: eligible dummy only; B: eligible dummy and non-eligible in treatment village dummy. C: model estimated on control sample only. Asymptotic standard errors in italics.

Estimation Results II

Figure: Parameter estimates for the initial condition model

	A	B	C
Poor	−0·275	−0·243	−0·280
	0·030	*0·046*	*0·051*
Ineligible individual in a PROGRESA village	—	0·057	—
	—	*0·055*	—
Father's education			
Primary	0·180	0·181	0·218
	0·025	*0·025*	*0·04262*
Secondary	0·262	0·264	0·281
	0·030	*0·030*	*0·05302*
Preparatoria	0·559	0·558	0·499
	0·0160	*0·057*	*0·09107*
Mother's education			
Primary	0·159	0·158	0·231
	0·026	*0·026*	*0·04446*
Secondary	0·316	0·314	0·398
	0·030	*0·030*	*0·05139*
Preparatoria	0·301	0·301	0·334
	0·061	*0·061*	*0·09740*
Indigenous	−0·005	0·006	0·133
	0·036	*0·026*	*0·0461*
Availability of Primary 1997	0·373	0·372	0·691
	0·073	*0·073*	*0·19003*
Availability of Secondary 1997	0·808	0·804	−0·568
	0·188	*0·188*	*0·349*
Kilometer to closest secondary school 97	0·00004	0·00004	−0·0002
	0·00024	*0·00003*	*0·00007*
Availability of Primary 1998	−0·261	−0·264	−0·449
	0·127	*0·126*	*0·235*
Availability of Secondary 1998	−0·845	−0·841	0·516
	0·187	*0·187*	*0·348*
Kilometer to closest secondary school 98	−0·0001	−0·0001	0·00015
	0·00003	*0·00003*	*0·00007*
Cost of attending secondary	0·00006	0·0001	−0·00019
	0·00024	*0·00024*	*0·00037*

Notes: As in Table 3. State dummies included. Availability means school in the village.

Estimation Results III

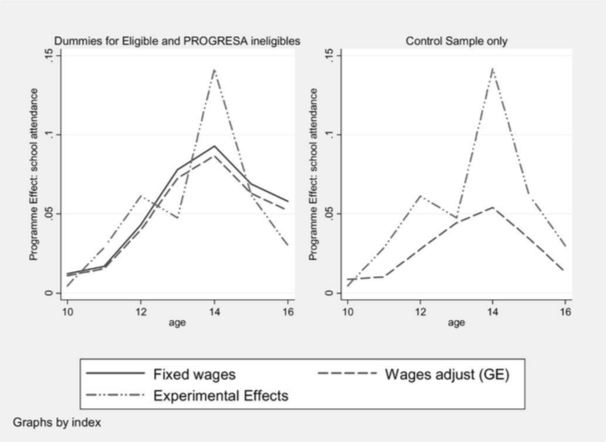
Figure: Parameter estimates for the education choice model

	A	B	C
Wage	0·134	0·168	0·357
	0·043	*0·045*	*0·100*
PROGRESA grant	3·334	2·794	—
	1·124	*0·796*	—

- The wage is expressed as a determinant of the utility of work (so given the positive coefficient, an increase in wages decrease school attendance)
- The grant coefficient is expressed as a ratio to the coefficient of the wage

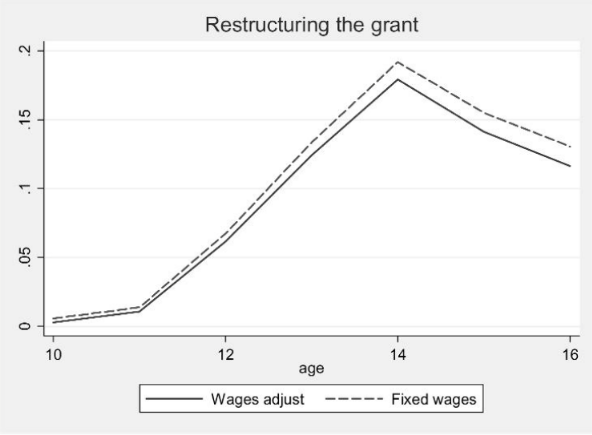
In-sample Fit of the estimated Model

Figure: Comparing treatment effects from the experiment with impacts based on the estimated model using both treatment and control villages (left panel) and only control villages (right panel)



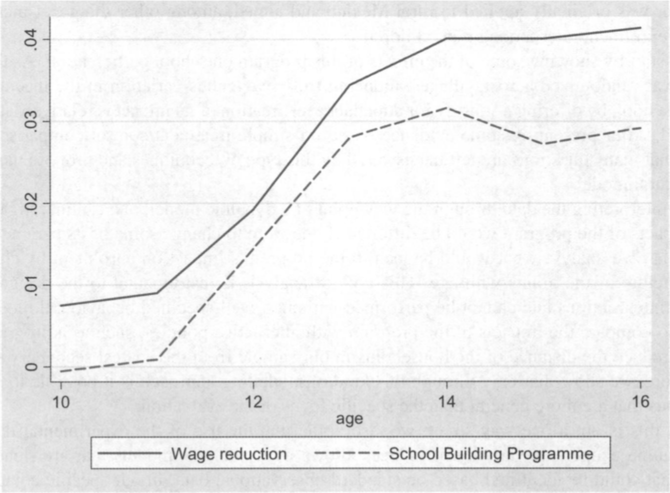
Counterfactual Policy Experiments I

Figure: Estimated program impact on enrollment rates under a budget neutral shift in the grant to those above grade 6 only



Counterfactual Policy Experiments II

Figure: Alternative policies to enhance school enrollment rates



Wrapping up on RCTs and DDC Models

If a model can provide a good forecast for a holdout sample that faces a policy regime well outside the support of the data (and that is not used in model formulation), then we should gain confidence that it can provide a good forecast of impacts of other policy changes along the same dimension

Figure: The Predicted Effect of Doubling the Subsidy for Children Aged 12-15

	Boys		Girls	
	$1 \times$ Subsidy	$2 \times$ Subsidy	$1 \times$ Subsidy	$2 \times$ Subsidy
Todd and Wolpin (2008): S-NP[1]				
Single child	0.056	0.116	0.060	0.141
Multiple children	0.059	0.078	0.070	0.089
Todd and Wolpin (2006): S-P	0.077	0.159	0.064	0.146
Attanasio, Meghir, and Santiago: S-P	0.070	0.131	na	na

Wrapping up on RCTs and DDC Models: Validation Vs. Identification?

- Should one have stronger belief in the predictions of the counterfactual experiments from TW as opposed to AMS because the former was externally validated?
    1. There exists a "true" model underlying the DGP that we observe (identification)
    2. There are models that perform better or worse than others in addressing particular questions (validation)