Empirical Methods for Policy Evaluation

Second Part

Matteo Bobba

Toulouse School of Economics (TSE)

TSE PhD Program (MRes)
Fall 2024

Ex-ante and Ex-post Policy Evaluation

Outline and Readings for this Class

1. Ex-ante policy evaluation
   - **Chapter 2 in Wolpin's book (MIT press, 2013)**
2. Combining causal inference and model-based analyses
   - Todd and Wolpin (AER, 2006)
   - Attanasio, Meghir, and Santiago (ReStud, 2012)

Ex-ante policy evaluation

Causal Inference Approach (only ex-post)

- Binary random variable $T_i = \{0, 1\}$ and potential outcomes (Y_i^1, Y_i^0), such that $Y_i = (1 - T_i)Y_i^0 + T_i Y_i^1$

$$
\begin{aligned}
ATE &= \mathbb{E}(Y_i^1 - Y_i^0) \\
&= \mathbb{E}(Y_i^1 | T_i = 1) - \mathbb{E}(Y_i^0 | T_i = 0) \\
&= \underbrace{\mathbb{E}(Y_i^1 - Y_i^0 | T_i = 1)}_{\text{ATT}} + \underbrace{\mathbb{E}(Y_i^0 | T_i = 1) - \mathbb{E}(Y_i^0 | T_i = 0)}_{\text{Selection bias}}
\end{aligned}
$$

- Research designs attempt to kill selection bias by means of identification assumptions. E.g.:

RCT	SUTVA
Diff-in-Diff	Parallel trends + SUTVA
IV-LATE	First-stage + exclusion restriction + Monotonicity + SUTVA
RD	Continuity (Sharp RD) + IV-LATE assumptions (Fuzzy RD) + SUTVA

Model-Based Approach (both ex-ante and ex-post)

1. Lay out an economic model of the phenomenon being studied
2. Addition of a stochastic structure if the model itself does not possess one
3. A consideration of the identification of the "primitive" model parameters given the data, model, and estimator employed

parametric: Show that two sets of parameters yielding the same likelihood value are necessarily equal

non-param: Show that the distribution of observables picks only one set of parameters irrespectively of the stochastic assumption on the distribution of unobservables

4. Adaptation of an estimation technique given the nature of the model and the data at hand
5. Given estimates of primitive parameters, empirical comparative statics exercises and/or counterfactual policy experiments

Ex-ante Policy Evaluation

- Economic models allow predicting the effects of public policies **before they are implemented** and/or variants of existing policies
  - Improve program design to maximize impacts given costs
  - Inform program targeting by identifying sub-populations for which impacts are highest
  - Analyze program impacts over a time horizon that exceeds the length of time observed in the data
  - Analyze program impacts in the presence of spillover or general equilibrium effects

An Example

- Many governments have adopted conditional cash transfer (CCT) programs as a way to alleviate poverty and stimulate human capital investments
  - Provide cash transfers to HHs conditional on school attendance of children
- Can we evaluate those programs before they are implemented?
  - Yes, with a model of schooling decisions in which the transfer decreases schooling costs

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

Economic Model

- Consider the following static optimization problem for the household

$$\max_{s\in\{0,1\}} U(c,s) \text{ s.t. } \begin{cases} c = y + w(1-s) \\ c = y + w(1-s) + \tau s \end{cases}$$

- Optimal schooling choices without and with the subsidy:

$$s^{\star} = g(y,w)$$
$$s^{\star\star} = g(\tilde{y},\tilde{w})$$

  - $\tilde{y} = y + \tau$ and $\tilde{w} = w - \tau$

- The impact of the subsidy is equivalent to a (income-compensated) reduction in child wages

Bringing the Model to the Data

- Add observables and unobservables preference shifters

$$U(c, s, X, \epsilon)$$

- Unobserved heterogeneity is not systematically related to wages and income

$$f(\epsilon | y, w, X) = f(\epsilon | X) \qquad \text{(CIA)}$$

- Given CIA, variations in wages and income identify the impact of the program

Non-Parametric Estimation

- Ex-ante average treatment effect is:

$$\hat{\Delta}_{np} = \frac{1}{N}\sum_{j=1}^{N}\left[\underbrace{\hat{\mathbb{E}}(s_i|w_i = w_j - \tau_j, y_i = y_j + \tau_j, X_i)}_{\text{Predicted schooling under the program}} - \underbrace{s_j(w_j, y_j, X_j)}_{\text{Observed schooling}}\right]$$

  - $\mathbb{E}(s_i|w_i = w_j - \tau_j, y_i = y_j + \tau_j, X_i)$ can be estimated by nonparametric regression (kernel, local linear regression or series estimation)
  - Need common support in the data: i.e. set of families with X_i for which the values $w_j - \tau$ and $y_j + \tau$ lie within the observed support of w_i and y_i

Counterfactual Subsidy Levels

	Boys		
Ages	2* Original	Original	0.75*Original
12-13	0.04 (59%)	0.01 (87%)	0.003 (98%)
14-15	0.24 (45%)	0.01 (83%)	0.05 (98%)
12-15	0.12 (53%)	0.06 (86%)	0.02 (98%)
	Girls		
	2* Original	Original	0.75*Original
12-13	0.06 (48%)	0.06 (91%)	0.05 (98%)
14-15	0.23 (51%)	0.07 (89%)	0.03 (98%)
12-15	0.14 (50%)	0.06 (90%)	0.05 (98%)
	Boys and Girls		
	2* Original	Original	0.75*Original
12-13	0.05 (54%)	0.04* (89%)	0.03 (98%)
14-15	0.23 (48%)	0.09 (86%)	0.04 (98%)
12-15	0.13 (52%)	0.06 (88%)	0.03 (98%)

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

Unconditional Income Grant

	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.

Adding Home Production

- Suppose that now we modify the model to allow for an alternative use of children's time, home production $l \in \{0, 1\}$

$$\max_{(s,l)} U(c, l, s) \text{ s.t. } \begin{cases} c = y + w(1 - s - l) \\ c = y + w(1 - s - l) + \tau s \end{cases}$$

- Optimal schooling choices without and with the subsidy are different

$$s^\star = g(y, w)$$
$$s^{\star\star} = h(\tilde{y}, \tilde{w}, \tau)$$

- Non-parametric ex-ante approach is not feasible
- Which policy restores the equivalence between the schooling demand functions?

Parametric Approach

- Consider the following functional form for the utility function under the original problem (for simplicity, no child leisure and no X)

$$U(C, s; \epsilon) = C + \alpha s + \beta C s + \epsilon s,\ \epsilon \sim N(0, \sigma_\epsilon^2)$$

- The probability of school attendance under the subsidy is

$$P(s = 1) = 1 - \Phi\left(\frac{(w - \tau) - \alpha - \beta(y + \tau)}{\sigma_\epsilon}\right)$$

- Model parameters can be estimated by ML from data with no subsidy ($\tau = 0$) given the same sources of variation mentioned in the non-parametric case

Parametric Approach

- Given parameter estimates, the effect of introducing a subsidy of τ on the attendance rate can be calculated from

$$\hat{\Delta}_p = \Phi\left(\frac{(w-\tau) - \hat{\alpha} - \hat{\beta}(y+\tau)}{\hat{\sigma}_\epsilon}\right) - \Phi\left(\frac{w - \hat{\alpha} - \hat{\beta}y}{\hat{\sigma}_\epsilon}\right)$$

- Unlike the non-parametric case, there is no condition on the support of $y+\tau$ and $w-\tau$
- Functional forms and distributional assumptions substantially decrease the computational burden (curse of dimensionality) in solving/estimating structural models

Static Vs. Dynamic Model to Evaluate a Cash Transfer Program

- 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 age 17)
- See Todd and Wolpin (2006) and Attanasio, Meghir and Santiago (2012)

Wrapping Up on Ex-Ante Policy Evaluation

- Estimating the effect of a new policy does not necessarily require specifying the complete structure of the model governing decisions
- Nonparametric ex ante policy evaluation may be feasible even when there is no variation in the data in the policy instrument (here, the price of schooling)
- If not feasible, one needs to impose extra-assumptions on the distribution of observed and unobserved heterogeneity

Combining Causal Inference and Model-Based Approaches

Out of Sample Validation

- Concerns about the plausibility of the model assumptions undermine the credibility of its predictions
  - Within-sample goodness-of-fit tests provide useful but not necessarily compelling evidence of the validity of the model
- The reliability of the model's predictions is better assessed in terms of **out-of-sample fit**
  - Estimate a model by holding out the treatment/control group, and then validate its predictions about program impacts

Out of Sample Validation – Todd and Wolpin (2006)

	Boys			
Ages	Experimental	Predicted	Sample-Sizes‡	% overlapping support
12-13	0.05** (0.02)	0.01 (0.03)	374, 10	87%
14-15	0.02 (0.03)	0.01* (0.04)	309. 569	83%
12-15	0.03 (0.02)	0.06 (0.03)**	683, 1179	86%
	Girls			
	Experimental	Predicted	Sample-Sizes‡	% overlapping support
12-13	0.07 (0.07)	0.06* (0.03)	361, 589	91%
14-15	0.11** (0.04)	0.07 (0.05)	316, 589	89%
12-15	0.09 ** (0.02)	0.06** (0.03)	677, 1180	90%
	Boys and Girls			
	Experimental	Predicted	Sample-Sizes‡	% overlapping support
12-13	0.06** (0.02)	0.04* (0.02)	735, 1199	89%
14-15	0.07** (0.03)	0.09** (0.04)	625, 1160	86%
12-15	0.06** (0.02)	0.06** (0.02)	1360, 2359	88%

†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.

Identification and Estimation

- One can directly use the source of variation induced by the program for **estimation** of the model parameters
  - Estimate a model on both treatment and control groups that relaxes some behavioral/distributional assumptions
- In the previous model, the impact of the subsidy on schooling is assumed equivalent to a decrease in child wages
  - Transfers are actually handed out to the mother, while we do not know who receives the child's wage
  - Who receives the money likely matters

Identification and Estimation – Attanasio et al (2012)

- Consider the alternative model:

$$U^s - U^w = \alpha + (\beta^s - \beta^w)Y + \theta^s \tau - \theta^w w$$

- Previous model assumes income pooling conditional on schooling ($\theta^s = \beta^s$ and $\theta^w = \beta^w$)
- By estimating on the control group only, $\tau = 0$ and so we have to impose that the transfer and the wage have the same effect on schooling decision ($\theta^s = \theta^w$)
- Estimating the model on both treated and control villages enables to identify both θ^s (through variations in transfer) and θ^w (through variation in child wages)

Validation Vs. Identification?

- Should one have stronger belief in the predictions of the counterfactual experiments from Todd and Wolpin (2006) as opposed to Attanasio et al (2012) because the former was externally validated?
    - Attanasio et al (2012) may be more credible for being parsimonious, and yet more general!
- It is difficult to account for all the possible behavioral responses in a model estimated off the control group only
    - Use all the data at your disposal

Practical Considerations for Bridging the two Approaches

1. Show your data/variation with descriptive analysis
2. Use the design-based analysis to provide preliminary evidence
3. Clearly articulate the value-added of the model
4. Use the design-based analysis to guide modeling choices and identification
5. Choose parameters of interest and counterfactuals that are directly informed by the variation in the data