Structural models and experimental methods: complements or substitutes?

Orazio P. Attanasio

UCL, IFS, NBER & BREAD
o.attanasio@ucl.ac.uk

PODER SUMMER SCHOOL at the Paris School of Economics
July 1-2 2015
Lecture 1
Using an RCT to identify a structural model
Lecture 1
Using an RCT to identify a structural model

Endogeneity and causality in economics: the case for Randomized Controlled Trials.

What can and what can’t be learned from a RCT.

The PROGRESA evaluation: Conditional Cash Transfers and School Enrollment.

i. Using an RCT to validate a structural model.
ii. Using a structural model to extrapolate the results of a RCT.

Technical Digression I. Different types of randomizations:

i. Individual level randomisation.
ii. Clustered randomization.
Lecture 1
Using an RCT to identify a structural model
....and using the structural model to extrapolate the RCT results

1. Endogeneity and causality in economics: the case for Randomized Controlled Trials.
2. What can and what cant be learned from a RCT.
3. The PROGRESA evaluation: Conditional Cash Transfers and School Enrollment.
   i. Using an RCT to validate a structural model.
   ii. Using a structural model to extrapolate the results of a RCT.
4. Technical Digression I. Different types of randomizations:
   i. Individual level randomisation.
   ii. Clustered randomization.
Lecture 2
Using a structural model to understand the impact of an RCT.

1. The experiment and its impacts.
2. A model of HK accumulation.
3. Estimating part of the model: what variation?
4. Interpreting the results of the experiment.
5. Technical Digression II. Multiple hypothesis testing.
6. Technical Digression III. Eliciting subjective expectations
7. Eliciting beliefs
Impact evaluation: causation and endogeneity

- The problem of establishing causation from observations is a very old problem.
The problem of establishing causation from observations is a very old problem.

It has been discussed extensively in statistics, philosophy of science, epidemiology, social sciences.

- The design of experiments.
  - Fischer, Neyman, Rubin.
- The causes of cholera.
  - John Snow and the Broad Street Water Pump (1854)
- The impact of job training programs.
  - The Ashenfelter dip.

Selection, participation choices and endogeneity:
- The returns to education.
The problem of establishing causation from observations is a very old problem.

It has been discussed extensively in statistics, philosophy of science, epidemiology, social sciences.

- The design of experiments.
  - Fischer, Neyman, Rubin.
- The causes of cholera.
  - John Snow and the Broad Street Water Pump (1854)
- The impact of job training programs.
  - The Ashenfelter dip.

Selection, participation choices and endogeneity:
- The returns to education.
- Comparative advantage and the Roy model.
Counterfactuals

- Suppose we are interested in learning the impact of a treatment $D$ on an outcome $Y$.
- We conceptualise this problem assuming that, for each individual $i$, there are two potential outcomes:

$$Y_i(D = 1)$$

and

$$Y_i(D = 0)$$
Suppose we are interested in learning the impact of a treatment $D$ on an outcome $Y$.

We conceptualise this problem assuming that, for each individual $i$, there are two potential outcomes:

$Y_i(D = 1)$

and

$Y_i(D = 0)$

For each individual it is logically impossible to observe both outcomes.

However, if we have a large number of individuals, we can estimate the average impact.
Counterfactuals

- Suppose we are interested in learning the impact of a treatment $D$ on an outcome $Y$.
- We conceptualise this problem assuming that, for each individual $i$, there are two potential outcomes:

$$Y_i(D = 1)$$

and

$$Y_i(D = 0)$$

- For each individual it is logically impossible to observe both outcomes.
- However, if we have a large number of individuals, we can estimate the average impact.
- This is easy if there is no relationship between the outcome of interest and the assignment of the experiment.
Counterfactuals

- Suppose we are interested in learning the impact of a treatment $D$ on an outcome $Y$.
- We conceptualise this problem assuming that, for each individual $i$, there are two potential outcomes:
  \[ Y_i(D = 1) \]
  and
  \[ Y_i(D = 0) \]
- For each individual it is logically impossible to observe both outcomes.
- However, if we have a large number of individuals, we can estimate the average impact.
- This is easy if there is no relationship between the outcome of interest and the assignment of the experiment.
- This is the case if the assignment is random, as in an experiment.
Outcomes are heterogeneous across individuals.

Heterogeneity is driven by many factors, some observables, many unobservable.

In general, heterogeneity also drives selection into a programme.

Indeed individuals will select into intervention on the basis of comparative advantage.
Experiments and quasi-experiments

- Experiments might be difficult to design and implement for many considerations.
- At times quasi-experiments are available that can be used for inference.
- This is the logic behind diff-in diff approaches.
Experiments and quasi-experiments

Example: John Snow “On the Mode of Communication of Cholera” (1854):

No fewer than three hundred thousand people of both sexes, of every age and occupation, and of every rank and station, from gentle folks down to the very poor, were divided into two groups without their choice, and, in most cases, without their knowledge; one group being supplied with water containing the sewage of London, and amongst it, whatever might have come from the cholera patients—the other group having water quite free from such impurity.

- Two water companies serving the same street, one which changed water source in 1852, a year before a cholera epedemics in London.
John Snow evidence and the explanation of the cholera’s epidemics

- Despite the evidence collected by John Snow (on the two water companies and then the Broad Street Pump) the germ theory of propagation of cholera was not accepted.
- In 1854 the anatomist Filippo Pacini identified the micro-organism that transmits cholera.
- At that point the ‘model’ could be matched to empirical evidence and became eventually accepted.
What are the advantages of RCTs’?

- They provide very strong evidence.
- They are ‘theory’ free.
- They are easily explained to policy makers.
What are the risks of RCTs?

- Deviations from SUTVA.
  - Stable Unit Treatment Value Assumption
- Hidden selection (attrition?)
- Non-compliance.
What are the limitations of RCTs?

- External validity: difficulty in extrapolating.
- General equilibrium effects
- Is the estimated parameter the one that is relevant for policy?
Can structural models complement Randomized Controlled trials?

- Structural models impose a structure on the data and, as such,
- The variability induced by a randomised controlled trial is by definition exogenous.
- There are two ways these two approaches can interact:
  - We can use experimental information to validate a model (Todd and Wolpin).
  - We can use the same information to help identify a possibly richer structural model (Attanasio, Meghir, Santiago).
But why use the structural model in this context if we already "know" the answer from our experiments?

The structural model will help us interpret the data and understand the mechanisms through which an intervention works.

The model may allow simulation of alternative policies thereby offering a mechanism for improving effectiveness.

Finally, validation offers the possibility of understanding better the shortcomings of models.
The PROGRESA trial

- In 1997 the Mexican government designed a new intervention to combat long term poverty in marginalised rural communities.
- PROGRESA was one of the first conditional cash transfers.
- It was initially targeted at about 10,000 marginalized rural localities.
- It subsequently expanded to the entire country:
  - It currently covers over 10% of Mexican population
  - It is the largest welfare programme in Mexico
  - Its name was changed first to *Oportunidades* and, recently, to *Prospera*. 
Targeting was first done at the locality level.

Once a locality was targeted a census was conducted to determine the eligible households.

- Eligibility was determined by the presence of children younger than 17 and by a wealth index.
- About 75% of households would receive the programme.

Eligible households were entitled to a cash transfer if they complied with certain conditions:

- If they had children 6 they had to take them periodically to health centres.
- If they had school age children they had to enroll and attend school regularly.
- Mothers had to attend classes and other activities.
- Grants were larger for higher grades and for girls.
- Transfers were targeted to women.

The grant is substantial (about 20-25% of income).
The expansion of the first phase of the programme lasted 2 years.

The timing of the expansion used to design an evaluation based on a Clustered Randomized Controlled Trial.

In 1997, the Mexican government identified 506 villages:
- 320 were randomized for an early start (in April 1998)
- 286 were randomized for a late start (December 1999)

Extensive surveys (covering all households in the 506 villages) were collected in:
- March and October 1998
- March and November 1999
- April 2000, March 2003, March 2007
PROGRESA: the impacts.

- The programme had positive impact on:
  - Nutrition and growth
  - Consumption and poverty
  - Secondary school enrolment (5-6% from 67%)
- No impact on primary school enrolment (already over 90%).
- Some increases in the consumption of non-beneficiaries.
  - Transfers? Spillover effects.
- Small increase in children wages:
  - GE effects
- Notice that the evaluation design allows the identification of GE and spillover effects
- Social capital? Women status?
The impacts are easy to compute:
- Compare treatment and control villages
  (the randomization worked).

But what are the mechanisms?

Let's take schooling decision:
- The program changes the relative price of schooling.
- The program also has an income effect

We can model the decision of sending children to school.
Parents decide *in the child's best interest* whether the child works or attends school.

We assume households decide on children school enrolment taking into account:
- Future returns to education (appropriately discounted);
- Cost and availability of schools (proxied by distance to schools, direct costs etc.);
- The opportunity cost of school (lost income) information on children wages;
- The grant (in the treatment villages);
- Other variables.

We can then estimate this model and its parameters.

The presence of the RCT allows us the identification of a rich model.

We can then simulate test the model fit and simulate it.
The flow utility is defined by:

\[ u_{it}^s = Y_{it}^s + \alpha g_{it} \]
\[ Y_{it}^s = \mu_i^s + a^s z_{it} + b^s d_{it} + 1(p_{it} = 1)\beta^p x_{it}^p + 1(s_{it} = 1)\beta^s x_{it}^s + \epsilon_{it}^s \]
Model overview

- In the above $u_{it}^s$ is the current utility of going to school.
- This depends on current costs $Y_{it}^s$ and on the grant ($g_{it}$ for it those eligible, zero for the others)
- The current costs depend on unobserved ability $\mu_{it}^s$, on costs of attending primary ($x_{it}^p$) or secondary ($x_{it}^s$) education and on household characteristics as well as on a random shock.
- The utility from school also depends on accumulated schooling, the idea being that going to school may actually strengthen attachment.
Model overview

- The utility from work is simply:

  \[ u_{it}^w = \delta w_{it} \]

  where \( w_{it} \) represents the market wage that children can earn. The wage is thus the opportunity cost of schooling.
The utility from work is simply:

\[ u_{it}^w = \delta w_{it} \]

where \( w_{it} \) represents the market wage that children can earn. The wage is thus the opportunity cost of schooling.

Parents are assumed to maximise the present discounted value of the utility flow from the current age to age 18 + the terminal value function that is a function of the level of education achieved by the child.
Why is this problem dynamic?

- Education has benefits in the future
- Past education can change attitudes towards attendance
- The grant itself creates dynamics because not going to school one year reduces the total number of years the child can be subsidised: the grant is only available until 17.
Model overview

- We assume that parents choose education to maximise lifetime utility starting age 9.
- Decisions are taken from age 9 to 18. (Before that nearly all attend.)
- Terminal Value Function
  - At 18 adult life starts with a value of $V(ed_{i,18})$.
  - This defines in a reduced form way what the accumulated education is worth and needs to be estimated.
  - In a model where we follow people up later in life, the terminal value would be pinned down by labour market outcomes.
- We specify

$$V(ed_{i,18}) = \frac{\alpha_1}{1 + \exp(-\alpha_2 ed_{i,18})}$$
There are two sources of uncertainty:

- The random shocks to the cost of schooling
- The possibility that the child will not pass the grade pts. This depends on grade and age and is known to all concerned.

With richer data this probability could be made to depend on effort, thus making it endogenous.
Laws of motion

- The law of motion for the state variable edit is:
  
  \[ ed_{it+1} = ed_{it} + 1, \quad \text{if attend and pass grade} \]
  
  \[ ed_{i,t+1} = ed_{it}; \quad \text{otherwise} \]

- The variables \( z_{it} \) have a deterministic path known to everybody (simplifying assumption)
The value function for attending school is $V_{it}^s(ed_{it}|z_{it})$

The value function for work is denoted $V_{it}^w(ed_{it}|z_{it})$

Thus the value of school at age $t$ can be written as:

$$
V_{it}^s(ed_{it}|z_{it}) = u_{it}^s + \beta p_t^s(ed_{it} + 1) \times \\
E\max \left[ V_{it+1}^s(ed_{it} + 1|z_{it+1}), V_{it+1}^w(ed_{it} + 1|z_{it+1}) \right] + (1 - p_{it}^s(ed_{it} + 1)) \times \\
E\max \left[ V_{it+1}^s(ed_{it}|z_{it+1}), V_{it+1}^w(ed_{it}|z_{it+1}) \right]
$$

And the value of working is:

$$
V_{it}^w(ed_{it}|z_{it}) = u_{it}^w + \beta E\max \left[ V_{it+1}^s(ed_{it}), V_{it+1}^w(ed_{it}) \right]
$$
Computing the Emax function

- In this problem the Emax functions can be easily computed analytically
- Denote
  \[ u_{i_t}^s = \tilde{u}_{i_t}^s + \epsilon_{i_t} \]
- Then we have that at any point in the lifecycle

\[
\begin{align*}
E_{\max}\{V_{it}^s(ed_{it}), V_{it}^w(ed_{it})\} &= \\
E_{\max}\{\tilde{u}_{i_t}^s + \epsilon_{i_t} + \beta V_{it+1}^s(ed_{it}), u_{it}^w + \beta V_{it+1}^w(ed_{it})\} &= \\
E(\epsilon_{i_t} | \epsilon_{i_t} > u_{it}^w - \tilde{u}_{i_t}^s + \beta [V_{it+1}^w(ed_{it}) - V_{it+1}^s(ed_{it})] \times P^S + [u_{it}^w + \beta V_{it+1}^w(ed_{it})] \times (1 - P^S))
\end{align*}
\]
Computing the Emax function

In the above

\[ P^S = Pr(\text{Attend}) \]
\[ = Pr(\epsilon_{it} > u^w_{it} - \tilde{u}^s_{it} + \beta[V^w_{it+1}(ed_{it}) - V^s_{it+1}(ed_{it})]) \]

The term \( E(\epsilon_{it}|\epsilon_{it} > u^w_{it} - \tilde{u}^s_{it} + \beta[V^w_{it+1}(ed_{it}) - V^s_{it+1}(ed_{it})]) \) has closed form expression for the logistic.

In this case the DP becomes computationally as easy as any nonlinear static regression.
We estimate an equation for predicting wages. We do this for several reasons:

- Child wages are likely to be measured with error. We use the village adult wage, observed everywhere as an instrument.
- Wages are not observed for non-working kids. We thus correct for selection the estimated wage equations and predict wages for non-working children.
- We want to test for General Equilibrium effects.
Strictly speaking we should be integrating out wages - not predicting them.

In a linear model the two are identical.

This is however a nonlinear model (because of the future value functions)

What we do is just a simplifying approximation.

The wage equation we obtain is

\[
\ln w_{ij} = -0.983 + 0.0605P_j + 0.883 + \ln w_{ag} + 0.066\text{age}_{ij} + 0.0116\text{educ}_{ij} - 0.056\text{Mills}_{ij} + \xi_{ij}
\]

Note that the wage equation does not depend on education for the children.

This is because we found that education has no wage returns in the village economy (perhaps 1% a year)

Returns to education are enjoyed by those who obtain it by migrating and working in urban centres in adult life.
A general equilibrium argument and the wage equation

- Suppose we could approximate the implied human capital supply to the labour market of children \((c)\) and adults \((a)\) by

\[
H_k = L_k w_k^\gamma, \quad k = c, a
\]

- Production is governed by

\[
Q = A\left[\delta H_c^\sigma + (1 - \delta) H_a^\sigma\right]^{\frac{1}{\sigma}}, \quad \sigma < 1
\]

- The first order conditions imply

\[
\frac{w_c}{w_a} = \frac{\delta}{1 - \delta} \left(\frac{H_c}{H_a}\right)^{\sigma-1}
\]

- From Labor supply we get that

\[
\frac{H_c}{H_a} = \frac{L_c w_c^{\gamma_c}}{L_a w_a^{\gamma_a}}
\]
A general equilibrium argument and the wage equation

- Why should the child wage equation depend on the adult agricultural wage?
  - This can be shown by solving for equilibrium in a market with two labour inputs, child and adult labour.
  - The solution to this is

\[
\ln w_c = \frac{\rho + \gamma_a}{\rho + \gamma_c} \ln w_a - \left[ \frac{1}{\rho + \gamma_c} \ln(L_c) - \ln(L_a) + \kappa \right], \quad \rho = \frac{1}{1 - \sigma}
\]

- \(\gamma_k\), \(k = c, a\) are the adult and child labour supply elasticity.
- \(L_k\), \(k = c, a\) are the level of labour supply in the village.
Consider an simple static model for education choice

\[ U^s = \beta^s Y + \theta^s g \]
\[ U^w = \beta^w Y + \theta^w - a \]

The utility gain from school is

\[ U^s - U^w = (\beta^s - \beta^w) Y + \theta^s + \theta^w + a \]

The grant \( g \) and the wage \( w \) are allowed to have different effects.
Effect may be different because of intra household allocation reasons: grant goes to mother.

We do not know to whom the wage is paid: possibly to the child or even to the father.

Variability in the wage may not measure the impact of the grant ($\theta^s \neq \theta^w$).

However we need separate variability in the income from school and income from work to measure this.

The experiment offers this opportunity.

The experiment offers exogenous variation in dimensions we may not have in observational data.

The model offers a way to interpret the experimental variation.
Initial conditions

- We now need to deal with an important but difficult issue.
- We do not observe children as they enter school.
- We observe a cross section of children who at the start of the social experiment have some level of education $ed_{it}$ and some age $t$.
- The data consist of level of schooling and whether the child attended school or not after the experiment started (as well as a wealth of other variables).
- We do not observe history of schooling.
- This level of education is endogenous because it is correlated with unobserved ability $\mu_i$.
- This is the initial conditions problem.
To understand the problem consider the probability of attendance as implied by the model above:

\[ P(\text{Attend}_it = 1|\text{zit}, x^p_{it}, x^s_{it}, \text{wage}_{it}, \text{ed}_{it}, \mu_i) \]

Since we do not observe \( \mu_i \) we need to integrate it out.

The joint distribution of attendance and \( \mu_i \) is

\[ G(\text{Attend}_it = 1, \mu_i|\text{zit}, x^p_{it}, x^s_{it}, \text{wage}_{it}, \text{ed}_{it}) \]

\[ = P(\text{Attend}_it = 1|\text{zit}, x^p_{it}, x^s_{it}, \text{wage}_{it}, \text{ed}_{it}, \mu_i) \]

\[ g(\mu_i|\text{zit}, x^p_{it}, x^s_{it}, \text{wage}_{it}, \text{ed}_{it}) \]
There are two potential channels through which past education (and the other characteristics) affect the probability of attendance:

i. Their causal effect on attendance
ii. Their correlation with the unobservable and hence with the ability composition of each education level

While we may be willing to assume that the characteristics and unobserved heterogeneity are independent this is impossible for education.

The entire path depends on ability: Higher and higher levels of education are associated with higher levels of ability.

Thus as we move from one education level to the next the ability composition changes.
Initial conditions

- The next difficulty is that we need to explain the stock of education we observe, with an instrument that does not affect current attendance.
- Current attendance depends on distance to school.
- We make past stock of schooling depend on the distance to school as it was in the past, relying for identification on new schools being built.
The likelihood contribution becomes

\[ L_i = \int_{\mu} P(A_i = 1 | X_{it}^{s,p}, w_{it}, ed_{it}, \mu_i) P(ed_{it} = e | X_{it}^{s,p}, w_{it}, ed_{it}, \mu_i) dg(\mu) \]

where \( X_{it}^{s,p} = \{ z_{it}, x_{it}^p, x_{it}^s \} \)

\( P(ed_{it} = e | z_{it}, x_{it}^{s,p}, w_{it}, ed_{it}, \mu_i) \) is a reduced form equation of the stock of education and it includes as an explanatory instrument distance from school in the past (\( dist_{it-1} \)).

We approximate \( g(\mu) \) with a discrete distribution, in this case just three points of support suffice.
For the initial condition model, we assume that the residuals of the stock of education are, conditionally on the unobserved heterogeneity, distributed as a normal. Therefore, conditional on \( \mu_i \), we have an ordered probit that is estimated jointly with the schooling decision model. We should stress that the wage we use in the estimation is the value predicted by equation (7) (with the exclusion of the Mills ratio). Such an equation accounts for endogenous selection and takes into account the effect that PROGRESA had on child wages, so that it imputes a higher value for treatment villages.

### 6. ESTIMATION RESULTS

In this section, we report the results we obtain estimating different versions of the dynamic programming model we discussed above. In particular, we will be discussing three different versions of the model. The first constitutes our basic model. In the second, we control for the pre-program difference in enrolment rates among non-eligible individuals in treatment and control villages with a dummy in the specification for schooling costs that identifies the group of non-eligible boys in treatment villages. Finally, we present the estimates obtained fitting a version of our model where we impose equality of the marginal utility of the wage and the grant as discussed in Section 4.1 following on from equation (5).

In Tables 3-5, we present estimates of the two versions of the basic model we mentioned above: the first column (A) of each table refers to the version that ignores differences in pre-program school enrolment between treatment and control villages, while in the second (B) they are accounted for by a dummy for non-eligible households in treatment localities. This dummy does not have a significant effect in the initial condition equation (Table 4) but is significant in the structural model of educational participation (Table 5). The two degree of freedom likelihood ratio test for excluding this variable has a \( p \) value of 0.8%. However, the parameters hardly change when we move between the two specifications and the substantive implications of the two models are the same.

### TABLE 3

*The distribution of unobserved heterogeneity*

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Point of Support 1</strong></td>
<td>-9.706</td>
<td>-8.327</td>
<td>-4.290</td>
</tr>
<tr>
<td></td>
<td>( 1.041 )</td>
<td>( 1.101 )</td>
<td>2.46</td>
</tr>
<tr>
<td><strong>Point of Support 2</strong></td>
<td>-14.466</td>
<td>-13.287</td>
<td>-17.62</td>
</tr>
<tr>
<td></td>
<td>( 1.173 )</td>
<td>( 1.208 )</td>
<td>3.144</td>
</tr>
<tr>
<td><strong>Point of Support 3</strong></td>
<td>-5.933</td>
<td>-4.301</td>
<td>-0.267</td>
</tr>
<tr>
<td></td>
<td>( 0.850 )</td>
<td>( 0.941 )</td>
<td>2.45</td>
</tr>
<tr>
<td><strong>Probability of 1</strong></td>
<td>0.513</td>
<td>0.518</td>
<td>0.490</td>
</tr>
<tr>
<td></td>
<td>( 0.024 )</td>
<td>( 0.023 )</td>
<td>0.032</td>
</tr>
<tr>
<td><strong>Probability of 2</strong></td>
<td>0.342</td>
<td>0.335</td>
<td>0.270</td>
</tr>
<tr>
<td></td>
<td>( 0.022 )</td>
<td>( 0.021 )</td>
<td>0.017</td>
</tr>
<tr>
<td><strong>Probability of 3</strong></td>
<td>0.145</td>
<td>0.147</td>
<td>0.240</td>
</tr>
<tr>
<td><strong>Load factor for initial condition</strong></td>
<td>0.108</td>
<td>0.102</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>( 0.016 )</td>
<td>( 0.014 )</td>
<td>0.013</td>
</tr>
</tbody>
</table>

*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.
### Main Results: Initial Conditions

<table>
<thead>
<tr>
<th></th>
<th>Column 1</th>
<th>Column 2</th>
<th>Column 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability of Primary 1997</td>
<td>0.373</td>
<td>0.372</td>
<td>0.691</td>
</tr>
<tr>
<td></td>
<td>0.073</td>
<td>0.073</td>
<td>0.19003</td>
</tr>
<tr>
<td>Availability of Secondary 1997</td>
<td>0.808</td>
<td>0.804</td>
<td>−0.568</td>
</tr>
<tr>
<td></td>
<td>0.188</td>
<td>0.188</td>
<td>0.349</td>
</tr>
<tr>
<td>Kilometer to closest secondary school 97</td>
<td>0.00004</td>
<td>0.00004</td>
<td>−0.0002</td>
</tr>
<tr>
<td></td>
<td>0.00024</td>
<td>0.00003</td>
<td>0.00007</td>
</tr>
<tr>
<td>Availability of Primary 1998</td>
<td>−0.261</td>
<td>−0.264</td>
<td>−0.449</td>
</tr>
<tr>
<td></td>
<td>0.127</td>
<td>0.126</td>
<td>0.235</td>
</tr>
<tr>
<td>Availability of Secondary 1998</td>
<td>−0.845</td>
<td>−0.841</td>
<td>0.516</td>
</tr>
<tr>
<td></td>
<td>0.187</td>
<td>0.187</td>
<td>0.348</td>
</tr>
<tr>
<td>Kilometer to closest secondary school 98</td>
<td>−0.0001</td>
<td>−0.0001</td>
<td>0.00015</td>
</tr>
<tr>
<td></td>
<td>0.00003</td>
<td>0.00003</td>
<td>0.00007</td>
</tr>
<tr>
<td>Cost of attending secondary</td>
<td>0.00006</td>
<td>0.0001</td>
<td>−0.00019</td>
</tr>
<tr>
<td></td>
<td>0.00024</td>
<td>0.00024</td>
<td>0.00037</td>
</tr>
</tbody>
</table>

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

The third column (C) in the tables presents estimates of the model obtained from the control sample only. In this case, the experiment is not used to estimate the model and all incentive effects are captured by the wage, which acts as the opportunity cost of education. The purpose of estimating it is to compare the predictions of a model estimated using the experiment to one that does not and relies on the equality of the marginal utility of the wage and the grant. For all specifications, the discount factor was estimated to be between 0.95 and 0.98. This value was obtained from a grid search over several values, for our favourite version of the model.

---

It turns out that approximately the same value of the discount factor maximizes the likelihood function both in Columns 1 and 2 of our tables. The standard errors we report are conditional on the value of the discount factor.
### TABLE 5
Parameter estimates for the education choice model

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wage</td>
<td>0.134</td>
<td>0.168</td>
<td>0.357</td>
</tr>
<tr>
<td></td>
<td>0.043</td>
<td>0.045</td>
<td>0.100</td>
</tr>
<tr>
<td>PROGRESA grant</td>
<td>3.334</td>
<td>2.794</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>1.124</td>
<td>0.796</td>
<td>—</td>
</tr>
</tbody>
</table>

We estimate all the versions of the model on the sample of boys older than 9 and younger than 17. All specifications include, both in the initial condition equation and in the cost of education equation, state dummies, whose estimates are not reported for the sake of brevity. In addition, we have variables reflecting parental education (the excluded groups are heads and...
Comparing the experimental effects to impacts estimated based on the model first using the experimental data (left panel) and second the control sample alone (right panel)

Figure 1
Comparing the experimental effects estimated based on the model first using the experimental data (left panel) and second the control sample alone (right panel)
We plot the result of this exercise in Figure 2, where we again show the results with no wage adjustment (continuous line) and with a GE adjustment. Performing the GE adjustment is now a bit more complicated than in the previous exercise. The amount by which children wages would change with the counterfactual grant structure has to be extrapolated. We do that by using the elasticities discussed in Section 4.5.

The graph shows that by targeting the grant to the older children we can almost double the impact relative to the predicted effect from the model shown in the left panel of Figure 1. This occurs with no effect on the school participation of the younger primary age children. This is not surprising since the grant hardly changes their behaviour in the first place because almost all children go to school below grade 6, making it an unconditional transfer for that age group.

The overall resources targeted to families with children do not change with this reform, but the incentive structure does. This change to the grant structure seems to suggest a modification to the program that would much improve its ability to increase enrolment rates. This is particularly important because the modified program costs, in the steady state, the same amount as the current one. From the point of view of the households, note that they receive the same amount of resources over time: what changes is when they receive them. If households can borrow against the future grant, then the only effect of this reform is to improve incentives for school participation at later ages. If, on the other hand, families are liquidity constrained, the trade-off may be more serious, particularly if the grant at a younger age affects nutrition or other child inputs.

Attanasio and Rubio-Codina (2008) show that the impact of the PROGRESA grant on a variety of nutritional outcomes for very young children does not depend on whether they have primary school age siblings. This might be an indication that a change in the grant structure as the one described might not have large negative effects. More recently, the Mexican government is piloting two versions of the program in which primary school grants are eliminated and secondary school grants are increased.

24. Attanasio and Kaufmann (2009) show that for the Oportunidades/PROGRESA population, liquidity constraints can be important.
We next consider two alternative experiments. In particular, we consider the effect of decreasing the wage by an amount equivalent to the grant and the effect of reducing the distance to school to a maximum of 3 kms. All three experiments are summarized in Figure 3. In all cases, we use the model B in the tables. No grant is our baseline.

First, we decrease the wage by an amount equivalent to the grant. We see that the effect of the wage is estimated to be much lower than the grant; for example at age 15 the incentive effect is less than half the one in Figure 1. This evidence re-emphasizes the point already made, that the experimental data provides information on behaviour that may not be available through standard observational data.

In the next experiment, we demonstrate the effects of a potential school building program that would reduce the distance of secondary schools to no more than 3 km. We consider this because it could constitute an alternative policy to subsidizing participation (although we do not claim that this policy is equivalent in terms of cost or in terms of other benefits such as better nutrition and its impact). According to our parameter estimates, the effect is modest as it would increase participation by just about 3 percentage points at age 15.

8. CONCLUSIONS

In this paper, we demonstrate the power of using an economic model to analyse data from a major social experiment, namely PROGRESA in Mexico. Conversely, we also show the usefulness of using experimental data to estimate a structural economic model. The welfare program we consider is an important example of how experimental data can be used to inform economic modelling.