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

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.

Impact evaluation: causation and endogeneity

- 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.

Impact evaluation: causation and endogeneity

- 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

$$Y_i(D = 0)$$

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.

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.

Heterogeneity in Social Sciences Settings

- 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 epidemics 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).

Randomised Experiments and Structural Model

- 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

- In1997 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*.

PROGRESA: some details

- 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
  - Tranfers were targeted to women
- The grant is substantial (about 20-25% of income)

PROGRESA: the evaluation.

- 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 Constrolled 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?

PROGRESA: what happened?

- The impacts are easy to compute:
  - Compare treatment and control villages
  - (the randomization worked).
- But what are the mechanisms?
- Lets 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.

PROGRESA: what happened?

- Parents decide *in the childs 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.

Model overview

- The flow utility is defined by:

$$
\begin{aligned}
u_{it}^{s} &= Y_{it}^{s} + \alpha g_{it} \\
Y_{it}^{s} &= \mu_{i}^{s} + a^{s} z_{it} + b^{s} ed_{it} + 1(p_{it} = 1)\beta^{p} x_{it}^{p} + 1(s_{it} = 1)\beta^{s} x_{it}^{s} + \epsilon_{it}^{s}
\end{aligned}
$$

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 μ_{i}^{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.

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.
- 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

Model overview

- 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

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

Uncertainty

- 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:

$$\begin{aligned} ed_{it+1} &= ed_{it} + 1, \quad \textit{if} \quad \textit{attend} \quad \textit{and} \quad \textit{pass} \quad \textit{grade} \\ ed_{i,t+1} &= ed_{it}; \qquad \textit{otherwise} \end{aligned}$$

- The variables z_{it} have a deterministic path known to everybody (simplifying assumption)

Value functions

- 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 :

$$
\begin{aligned}
V_{it}^{s}(ed_{it}|z_{it}) &= u_{it}^{s} + \beta p_{t}^{s}(ed_{it}+1) \times \\
& Emax\left[V_{it+1}^{s}(ed_{it}+1|z_{it+1}), V_{it+1}^{w}(ed_{it}+1|z_{it+1})\right] \\
& +(1-p_{t}^{s}(ed_{it}+1)) \times \\
& Emax\left[V_{it+1}^{s}(ed_{it}|z_{it+1}), V_{it+1}^{w}(ed_{it}|z_{it+1})\right]
\end{aligned}
$$

- And the value of working is:

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

Computing the Emax function

- In this problem the Emax functions can be easily computed analytically
- Denote

$$u_i^s t = \widetilde{u}_{it}^s + \epsilon_{it}$$

- Then we have that at any point in the lifecycle

$$\begin{aligned}
Emax\{V_{it}^s(ed_{it}), V_{it}^w(ed_{it})\} &= \\
Emax\{\widetilde{u}_{it}^s + \epsilon_{it} + \beta V_{it+1}^s(ed_{it}), u_{it}^w + \beta V_{it+1}^w(ed_{it})\} &= \\
\widetilde{u}_{it}^s + \beta V_{i,t+}^s(ed_{it}) + \\
E(\epsilon_{it}|\epsilon_{it} > u_{it}^w - \widetilde{u}_{it}^s + \beta \left[V_{it+1}^w(ed_{it}) - V_{it+1}^s(ed_{it})\right] \times P^S \\
+ \left[u_{it}^w + \beta V_{it+1}^w(ed_{it})\right] \times (1 - P^S)
\end{aligned}$$

Computing the Emax function

- In the above

$$\begin{aligned} P^S &= Pr(Attend) \\ &= Pr(\epsilon_{it} > u^w_{it} - \widetilde{u}^s_{it} + \beta[V^w_{it+1}(ed_{it}) - V^s_{it+1}(ed_{it})]) \end{aligned}$$

- The term $E(\epsilon_{it}|\epsilon_{it} > u^w_{it} - \widetilde{u}^s_{it} + \beta[V^w_{it+1}(ed_{it}) - V^s_{it+1}(ed_{it})])$ has closed form expression for the logistic.
- In this case the DP becomes computationally as easy as any nonlinear static regression.

Wages

- 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 everyhere 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

Wages

- 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

$$\begin{aligned} lnw_{ij} &= \underset{(0.384)}{-0.983} + \underset{(0.028)}{0.0605} P_j + \underset{(0.049)}{0.883} + lnw_j^{ag} + \underset{(0.027)}{0.066} age_{ij} \\ &+ \underset{(0.0065)}{0.0116} educ_{ij} - \underset{(0.053)}{0.056} Mills_{ij} + \xi_{ij} \end{aligned}$$

- 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 childen (c) and adults (a) by

$$H_k = L_k w_k^{\gamma_k}, \quad k = c, a$$

- Production is governed by

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

- 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

$$\frac{H_c}{H_a} = \frac{L_c}{L_a}\frac{w_c^{\gamma_c}}{w_a^{\gamma_a}}$$

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

$$lnw_c = \frac{\rho + \gamma_a}{\rho + \gamma_c} lnw_a - [\frac{1}{\rho + \gamma_c} ln(L_c) - ln(L_a) + \kappa, \quad \rho = \frac{1}{1 - \sigma}$$

- $\gamma_k, \quad k = c, a$ are the adult and child labour supply elasticity.
- $L_k, \quad k = c, a$ are the level of labour supply in the village.

RCT and structural model: complements?

- Consider an simple static model for education choice

$$\begin{aligned} U^s &= \beta^s Y + \theta^s g \\ U^w &= \beta^w Y + \theta^w - a \end{aligned}$$

- 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.

RCT and structural model: complements?

- Effect may be different because of intra household allocation reasons: grant goes to mother.
- We do not kow 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 oportunity.
- 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 μ_i
- This is the initial conditions problem.

Initial conditions

- To understand the problem consider the probability of attendance as implied by the model above:

$$P(Attend_{it} = 1 | zit, x^p_{it}, x^s_{it}, wage_{it}, ed_{it}, \mu_i)$$

- Since we do not observe μ_i we need to integrate it out.
- The joint distribution of attendance and μ_i is

$$G(Attend_{it} = 1, \mu_i | zit, x^p_{it}, x^s_{it}, wage_{it}, ed_{it})$$
$$= P(Attend_{it} = 1 | zit, x^p_{it}, x^s_{it}, wage_{it}, ed_{it}, \mu_i)$$
$$g(\mu_i | zit, x^p_{it}, x^s_{it}, wage_{it}, ed_{it})$$

Initial conditions

- 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 distnace to school as it was in the past, relying for identification on new schools being built.

Initial conditions

- 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.

Results: Unobserved Heterogeneity

TABLE 3
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.

Main Results: Initial Conditions

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*

Parameter Estimates

TABLE 5
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*	—

Simulations - Fitting the effect of the grant

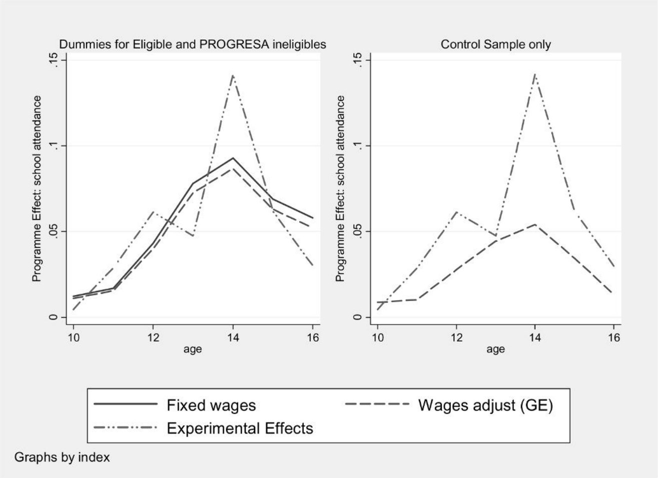

FIGURE 1
Comparing the experimental effects to impacts estimated based on the model first using the experimental data (left panel) and second the control sample alone (right panel)

Simulations - Restructuring the grant

REVIEW OF ECONOMIC STUDIES

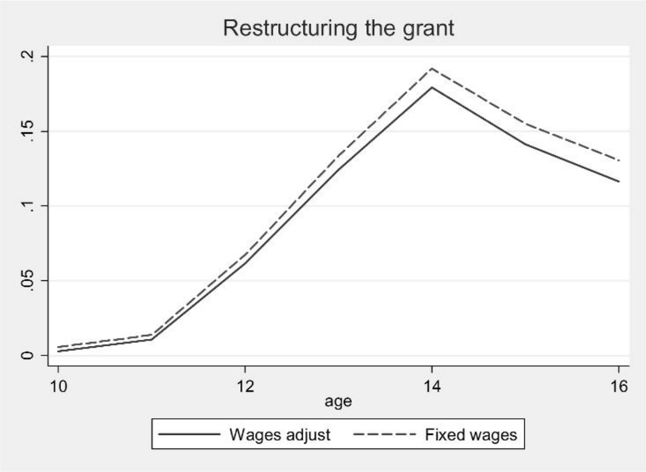

FIGURE 2

Redistributing the grant to those above grade 6 only—revenue neutral

Simulations - Alternative policies

ATTANASIO *ET AL.* EDUCATION CHOICES IN MEXICO

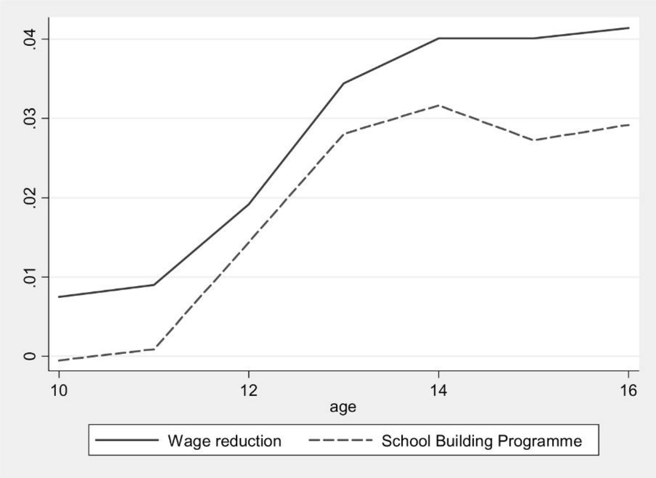

FIGURE 3
Alternative policy experiments