

**Comments on**

**“New Evidence on the Causal Link between the Quantity  
and Quality of Children” by J. Angrist, V. Lavy, and A. Schlosser**

**and**

**“Dynamic Models for Policy Evaluation” by Costas Meghir**

Manuel Arellano

Paris, December 15, 2005

# Comments on “New Evidence on the Causal Link between the Quantity and Quality of Children” by J. Angrist, V. Lavy, and A. Schlosser

Manuel Arellano

Paris, December 15, 2005

## 1. Summary

- A nice and thought provoking paper. Also hard and cleverly executed empirical work.
- It looks at the effect of number of siblings on the first-born child outcomes in families with at least 2 births (and also on the second child in families with at least 3 births).
- It belongs to a recent literature on the estimation of reduced form causal effects using twins or same-sex indicators as instrumental variables.
- Empirical approach as in Angrist–Evans, who looked at effects of family size on LS.
- In common with other papers, the negative OLS effect disappears in the IV estimates, suggesting that OLS estimates suffer from omitted variables bias.
- What are the distinctive features of this paper? A wider range of outcomes; exploiting ethnic diversity; a remarkable linkage of Israeli census and population registry.
- Looking at first-born and second-born subjects, together with the organization of the data into a 2+ and a 3+ sample, is nice because it helps defining meaningful effects.
- Thanks to the linked samples we are not restricted to youth still coresident with parents.

## 2. Data and variable construction

- AA status according to father's birthplace. What is the fraction of mixed (AA/non-AA) marriages? It appears that 20 % of the mothers of AA subjects are non-AA.
- Non-Jews (20% of Israeli population) not used because of lower census-to-population registry match rates. But if used, they would add further interesting diversity.

### *Adjusting the number-of-children variable for birth-spacing*

- For the sake of the argument, suppose it takes 15 years to reach adulthood. Consider a first-born subject who has 3 siblings (including the subject, i.e.  $c = 3$ ) and three cases:
  - Case 1: twins were born when the subject turned 15 so that sibship size was 1 for each of these 15 years ( $c^* = 1$ ).
  - Case 2: The subject was born as part of a triplet so that sibship size was 3 for each of the 15 years ( $c^* = 3$ ).
  - Case 3: One sibling born when subject was 5 and the other when 10. Sibship size was 1, 2, and 3 for one third of childhood each. Thus  $c^* = 1 \times \frac{1}{3} + 2 \times \frac{1}{3} + 3 \times \frac{1}{3} = 2$ .
- Alternative weighting schemes could be used if some periods of childhood are regarded as more important than others.
- OLS effects can be calculated much more generally, but IV effects depend on the number of instruments for identification.

### 3. Policy relevance

#### *Macro effects*

- Interest for policy is evaluation of family planning policies in developing countries.
- What do we learn from the paper's estimates in relation with this?
- Do family planning policies have macro effects that are not be captured by the present estimates? Connected to the spirit of the Indonesian coin: Family planning policies may help developing economies to pull away from the subsistence level equilibrium.
- In the Galor–Weil model, parents decide to have fewer children and send them to school in response to technological progress, which raises the return to human capital.
- The effectiveness of family planning policies depend on the existence on an externality (technical progress, crime rates), which is precisely what TEs cannot capture.
- In a developing country, which has imported health technologies but fewer production technologies, the quality–quantity relationship may be very different.

#### *Developed vs. developing economies*

- Non-western characteristics of the AA population? Not very compelling case for the relevance of AA results for developing societies.
- Caceres' point on sibship size effects on inputs that do not matter for outcomes (private school enrollment, shared-room) is nice, but less relevant for developing countries.

## *Specificities of Israeli case*

- Peculiarities of the Israeli population: large fraction of immigrants; threatened society; technologically advanced economy.

## *What causal effects?*

- The IV coefficients are causal in the sense that they can be interpreted as weighted averages of treatment effects for certain subpopulations. But can we stop at this point? Is it obvious that any average of treatment effects is worth attention?
- The set of controls used in the paper are dictated by what is available in the census and registry files. Are there controls that we would like to use but cannot because of data limitations (e.g. parents education or wealth)? If so what are the implications? It would be nice to add some discussion along these lines.
- The underlying model is some sort of production function for child outcomes: birth-spacing may matter; household resources will certainly matter.
- If birth-order matters effects for other siblings may be very different.

#### 4. Instrument validity

- The Rosenzweig-Wolpin (2000) critique: same-sex or having twins may have effects on the costs of children. Identification requires there is no such an effect.
- The birth of twins may have a direct effect on sibling outcomes aside from increasing family size through spacing.
- There is also the possibility that sex composition of siblings have an independent effect on children's outcomes.
- Butcher and Case (1994) use gender of siblings as an IV for schooling in an earnings equation. They argue that the gender of siblings may affect parental human capital investments in a given child through several channels, all of which go counter to using it as an instrument in a schooling equation:
  - differential preferences depending on the overall gender composition of children,
  - differences in the cost of raising girls relative to boys,
  - gender differences in returns to schooling.
- Black-Devereux-Salvanes find a positive and significant 2SLS effect of family size on educational outcomes for children when using same sex as instrument (but do not find it credible).

## 5. Discussion of results

### *OLS effects*

- The key sibship size margins are 2–3, 3–4, and 4–5. The OLS effects are negative and statistically significant but not very large: the effect of going from 2 to 5 siblings is less than half grade out of an average of 12.6 completed grades. Black et al. OLS results for Norway are even smaller, possibly related to using more controls.
- Are there nonlinearities in OLS effects? Why not replacing size by unrestricted indicators to check that? What are the  $R^2$ s? Nonlinear IV effects are harder to identify due to lack of variation in instruments (only dummy IVs).
- Basically, it would be nice to do a more detailed OLS analysis looking at nonlinearities, birth-order, spacing, etc.

### *IV effects*

- Table 6A reports positive, significant IV effects (for first borns in 3+Sample) on most schooling outcomes when using twins –this finding merits some comment.
- Apparently the indicator of twin second birth is not a significant predictor of family size for AA subjects: is there a weak instrument problem for the AA sample?

# Comments on “Dynamic Models for Policy Evaluation” by Costas Meghir

Manuel Arellano

Paris, December 15, 2005

## 1. Summary

- This paper illustrates the use of formal economic theory in policy evaluation research. It exhibits questions that can be answered with the help of theory but cannot with treatment effects approaches.
- It also points out the fragility of the answers to untested assumptions.
- This is a paper dealing with important, unresolved issues.
- Two topics are covered:
  - *Experiments as a source of structural identification*: Attanasio–Meghir–Santiago use experimental variation in costs of education to help identifying a structural model of education choices.
  - *GE effects*: Gallipoli–Meghir–Violante show that PE and GE effects of a tuition subsidy on college attendance *can* be very different (similar to Heckman, Lochner, and Taber, 1998).
- I just make some general points on these two themes.



## 2. Experiments and structural identification

- Experimental data have been used for estimating nonexperimental parameters in other contexts. Examples:
  - Intent to treat literature for biomedical randomized trials. Exogenous intent to treat but endogenous treatment status.
  - Estimation of the effect of training programs in duration models (Ham and LaLonde, 1996). Exogenous selection into treatment but endogenous selection into employment spells.
- In the first example one identifies a local average treatment effect.
- In the second one estimates a duration model with unobserved heterogeneity, which is nonstructural in the sense that it is not utility based. But the intention is to have a model with external validity i.e. that generalizes to a larger population or to other policies.
- In the Attanasio–Meghir–Santiago paper the objective is to use experimental variation to help the identification of a utility-based model.

### *Policy invariant parameters*

- In a structural model we assume that there exist effects which are constant regardless of the source of variation that induces the changes.
- Experimental data can be helpful in finding (or testing for) policy invariant parameters.
- Establishing policy invariance of structural parameters is crucial for credible structural policy evaluation (for external validity or valid extrapolation).
- Asserting that a parameter is structural because it appears in a utility formula does not make it more structural *per se*.
- Empirically relevant invariance is context specific. In the case of PROGRESA we would like to test for invariance to different stages of the program and to different subpopulations.

### 3. General equilibrium effects

*How long is the long run?*

- In the GE models used as illustrations only steady state equilibria are considered.
- The timing of the feedback effects may be of first-order importance for evaluation. Steady state effects may be less relevant if the adjustment process is long.
- Policy-evaluation strategies to estimating effects at various lengths from interventions may provide useful information about timing issues.

*At what level do general equilibrium effects take place?*

- If GE effects take place at the level of the economy (be a community, a region, or a country) we need randomization of economies to look at these type of effects, or observational studies that target such framework as a “standard”.
- Aggregate-level observational data (time series of economies), possibly in combination with micro data, are important in this respect.

## *Heroic assumptions*

- Imbalance between the detailed nature of the questions asked and the stylized assumptions used in the models.
- GE cannot be an excuse for assumptions that are known to be too restrictive in micro analyses of consumption, labor supply, or production.
- A word of caution: the excitement of 25 years ago about the estimation of the effect of taxes on labor supply using very elaborate budget sets together with very naive preferences that lead to incredible inferences.