# Comments to Context Matters for Size by L. Pritchett and J. Sandefur

S. Galiani
University of Maryland
IEN 2013

# Main Points of the Paper

- Lant and Justin (L&J) make a very important point. At this point, we cannot rely solely on experimental evidence to inform policy. I agree with this point.

- They also argue that we need to integrate all the evidence available. I also agree with that point. Less agreement might exists on how to do so and I have been struggling with the issue myself for years now.

- Finally, they argue that their evidence points in the direction of concluding that observational studies are, at least in some areas in the education literature, a better source of information, in an MSE sense, than the limited existent experimental evidence for policy advice.

# Main Points of the Paper

- This conclusion follows because of the existence of large effect heterogeneity, few experimental evidence, and presumably, low bias in observational studies of the effects of interest.

- Though this last point has merit, and it is true that there is much heterogeneity in program effects, and the experimental evidence is still limited, I am not persuaded that we know, or are often able to know, how biased are the estimates of observational studies.

- Clarification: Through the paper L&J refers always to OLS estimates (observational studies) versus experimental estimates (but sometimes they referred to quasi-experimental evidence).

# School Management

- Some years ago I did a survey on School Management and I was concerned with similar issues.

- We cover a non-exhaustive list of empirical papers that exploit non-experimental, quasi-experimental and experimental identification strategies.

- In our opinion, a deeper understanding of the structural mechanisms at work was a key goal for future research in this area.

- Given the heterogeneous effects of policies, knowing the channels through which they operate differentially across sub-populations or settings is of high priority.

# School Management

- We emphasized the question of comparison across results from different methodologies.

- For example, when we considered tracking, we reported observational or quasi-experimental studies that find no (or even a negative) effect of this intervention on students' performance, while we also reviewed randomized experiments showing a positive impact of tracking.

- The way in which we should reconcile the results is to understand the heterogeneity of both the settings and the methods exploited.

# School Management

- We said: “Unfortunately, for the time being, different settings are studied with different methodologies, so it is not obvious whether the differences in findings across studies are due only to differences in methodology or also to differences in the true parameters across the different settings.

- Even though experimental studies are internally valid, they do not necessarily have external validity (an issue obviously not exclusive to experimental studies, though). Therefore, the results cannot be generalized without further assumptions.”

- Thus, overall, in spirit, I share the concerns of L&J.

# Main Disagreements with L&J

- Seems to me that L&J assume that if a study is done in another population it lacks external validity but if it is from the population of interest it does not.

- But actually, both studies might (or might not) lack external validity.

- It also seems to me that L&J only consider the problem of omitted variables as the source of bias in observational studies. Not sure if this ends up being too relevant for their analysis, but conceptually, it is a narrow view of the problem.

# Main Disagreements

- My fundamental disagreement is about the estimation of the MSE in both cases.

- I find difficult to trust the estimates of the biases (which are not only due to omitted variables) of non-experimental estimators.

- And I also find difficult to see how to assess the variance of the true effect of an intervention across contexts (something that in my view is not only a problem of experimental setups).

# External Validity is very Important

- *Too many social scientists expect single experiments to settle issues once and for all. This may be a mistaken generalization from the history of great crucial experiments in physics and chemistry. In actually the significant experiments in the physical sciences are replicated thousands of times…. Because we social scientists have less ability to achieve "experimental isolation," because we have good reasons to expect our treatment effects to interact significantly with a wide variety of social factors many of which we have not yet mapped, we have much greater needs for replication experiments than do physical sciences.* **D. T. Campbell (1969)**

# Causal Generalization

- Related to external validity is the idea of causal generalization, which is concerned with specifying the range of application of a causal mechanism that has been identified with at least one instance of a treatment and outcome and at least one sample of persons and settings (Cook, 2000).

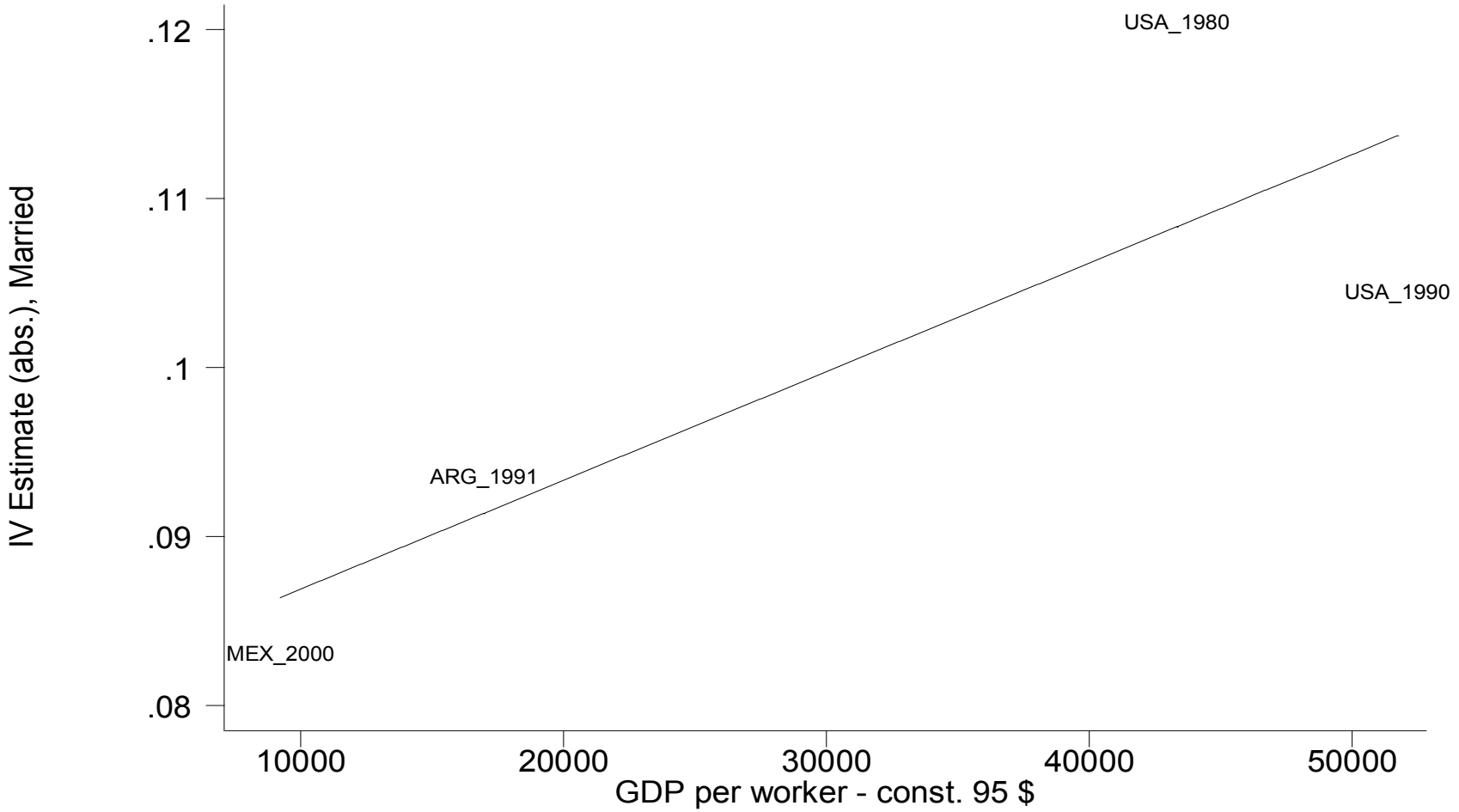
- In practice, there is a sense in which all causal generalization is about interpolation and extrapolation. Rubin (1992) suggests that causal generalization is about estimating a response surface, i.e., mapping a third variable to an estimated causal relationship. While this is clearly an advisable procedure in conceptual terms, it is difficult to attain in empirical work.

# Generalizing the Causal Effects of Fertility on Female Labor Supply

Guillermo Cruces and
Sebastian Galiani

Labour Economics 2007

**Figure 1. IV Coefficients and real GDP per worker**




Source: Angrist and Evans (1998) and author's estimations for the IV coefficients, corresponding to the "AE samples – married women" of the non-saturated model. The "GDP per worker (constant 1995 US dollars)" is based on the World Bank's (2002) series for constant GDP and workforce.