"Laboratory vs. Field Experiments: What Can We Learn?"

Comments on Levitt-List paper

by Al Roth

Roundtable discussion, ASSA meetings, Boston, January 8, 2006.

I like **both** lab and field experiments

In particular, I'm a great admirer of John's work.

Many points of agreement

- I certainly agree that we must be careful in interpreting and generalizing the results of experiments.
- But some of the terms of the discussion strike me as a bit skewed, and may obscure areas of broad agreement.
- For example, the titles…

“Laboratory **versus** field experiments”

- The title of our session seems like an odd way to talk about complementary goods.
  - Lab and field experiments are complements not only with each other, but also with other kinds of empirical and theoretical work

"What Do Laboratory Experiments Tell Us About the Real World?"

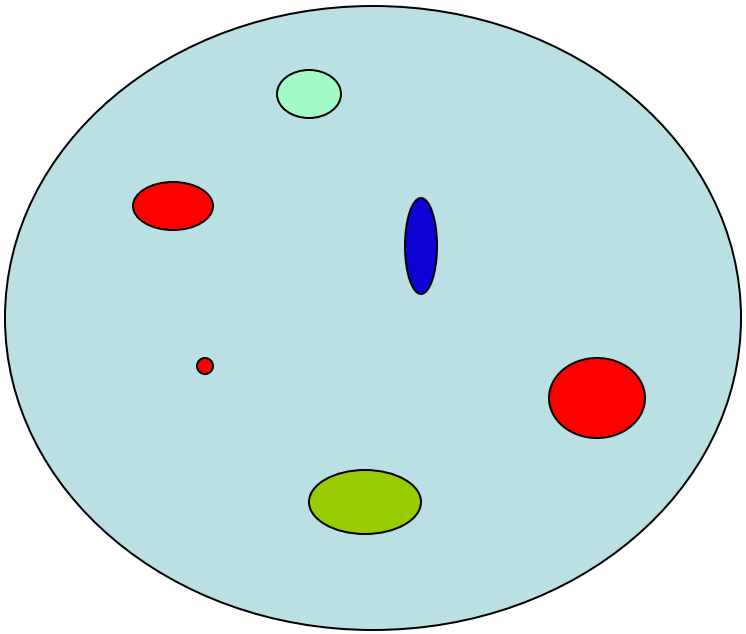
- I'm puzzled by what seems to be the implied *geography* of the Levitt-List paper.

Implicit geography of lab and world:

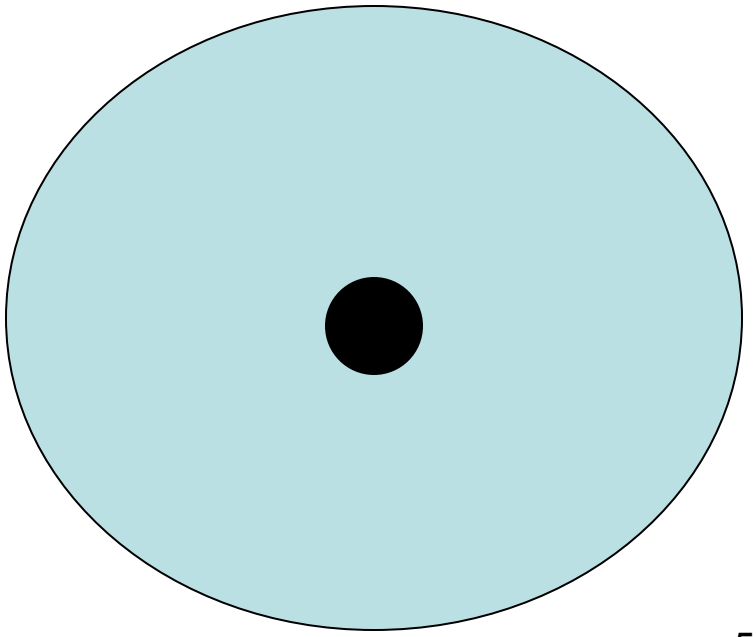
many points in lab space, only one in real world (and so no questions about robustness of observations except in the lab…)

- Lab experiments

(experimental conditions)



"Real World"



Recap:

- There are more generalizations to worry about than just from lab to real world
  - I fully agree that over-exuberant generalization is one of the things that most reduces the pleasure of reading an economics paper (whether the paper is a lab or field experiment, another kind of field study, or a theory paper…)
- I even agree that lab and field experiments may have different comparative advantage

Some things at which lab experiments excel

- Testing theoretical propositions
  - While controlling for the assumptions of the theory
- Abstracting away from the complexities of field data.
  - Whispering in the ears of princes

Some examples from market design

- Several of the clearinghouse markets I've helped design are similar from the point of view of the underlying economic theory
  - American and Canadian markets for medical residents (1998)
    - The same clearinghouse mechanism has subsequently been adopted by Postdoctoral Dental Residencies, Psychology Internships, Neuropsychology Residencies, Osteopathic Internships, Pharmacy Practice Residencies, Articling Positions with Law Firms in Alberta.
  - New York City high school matching (2003)
  - Boston public school matching (2006)
  - Gastroenterology match (2006)
- These different markets are all *different* points in the "real" world. How to draw some general conclusions about them?

Market	Stable	Still in use (halted unraveling)
• NRMP	yes	yes (new design in ’98)
• *Edinburgh ('69)*	*yes*	*yes*
• *Cardiff*	*yes*	*yes*
• *Birmingham*	*no*	*no*
• *Edinburgh ('67)*	*no*	*no*
• *Newcastle*	*no*	*no*
• Sheffield	no	no
• Cambridge	no	yes
• London Hospital	no	yes
• Medical Specialties	yes	yes (~30 markets, 1 failure)
• Canadian Lawyers	yes	yes (Alberta, no BC, Ontario)
• Dental Residencies	yes	yes (5) (no 2)
• Osteopaths (< '94)	no	no
• Osteopaths (≥ '94)	yes	yes
• Pharmacists	yes	yes
• Reform rabbis	yes (first used in ‘97-98)	yes
• Clinical psych	yes (first used in ‘99)	yes

So stability looks like an important feature in the success of a centralized labor market clearinghouse, based on diverse field observations.

The need for experiments

- How to know if the difference between stable and unstable matching mechanisms is the key to success?
  - There are other differences between e.g. Edinburgh and Newcastle
  - There are even bigger differences between British medical labor markets and
    - American medical and other health care markets
    - School matching

Simple laboratory experiments help show stability of the algorithm is critical

- Kagel, John H. and A.E. Roth, "The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment," *Quarterly Journal of Economics*, February, 2000, 201-35.)
  - Compares stable versus unstable mechanisms (those used in Edinburgh and Newcastle) in a market with no other differences (and just 6 subjects on each side of the market)
  - The virtue of the experiment is not that it reproduces the markets in Edinburgh and Newcastle, let alone the American medical or other markets.
  - Instead, it gives a simple, transparent comparison in which the only variable is the clearinghouse mechanism.
    - So it clarifies the effect of the mechanism in a way none of the field comparisons could.

Market	Stable	Still in use (halted unraveling)
• NRMP	yes	yes (new design in ’98)
• *Edinburgh ('69) yes*		*yes*
• *Cardiff*	*yes*	*yes*
• *Birmingham*	*no*	*no*
• *Edinburgh ('67) no*		*no*
• *Newcastle*	*no*	*no*
• Sheffield	no	no
• Cambridge	no	yes
• London Hospital no		yes
• Medical Specialties	yes	yes (~30 markets, 1 failure)
• Canadian Lawyers	yes	yes (Alberta, no BC, Ontario)
• Dental Residencies	yes	yes (5) (no 2)
• Osteopaths (< '94)	no	no
• Osteopaths (≥ '94)	yes	yes
• Pharmacists	yes	yes
• Reform rabbis	yes (first used in ‘97-98)	yes
• Clinical psych	yes (first used in ‘99)	yes
• Lab experiments	yes	yes.
	no	no

Lab experiments fit nicely on the list, just more of a variety of observations that increase our confidence in the robustness of our conclusions, the lab observations are the smallest but most controlled of the markets on the list…

Lab experiments can be tailored to help generalize other complex phenomena observed in the field

- For the gastroenterology market:
  - C. Nicholas McKinney, Muriel Niederle, and Alvin E. Roth, "The collapse of a medical labor clearinghouse (and **why such failures are rare**)," *American Economic Review*, *95, 3, June, 2005,878-889.*
- For schools:
  - Y. Chen and T. Sönmez "School Choice: An Experimental Study," *Journal of Economic Theory*, forthcoming.
- For decentralized markets:
  - Niederle, Muriel and Alvin E. Roth, "Market Culture: How Norms Governing Exploding Offers Affect Market Performance," working paper.
- *These experiments don't simply elicit behavior that generalizes to natural markets, but they do it in environments so simple and controlled that **causality** can be inferred, from within-experiment comparisons.*

In conclusion

- Broad agreement:
  - generalizations must be made carefully
    - From experiments and from field observations
- Field and laboratory experiments both add to our ability to understand the ("real") world
- ***Series*** of experiments, and ***varieties*** of observations help us understand what is robustly generalizable.

Alternative geography of the "real world"

(generalization is no less an issue in the field)