Use of Experimental Methods

(G&G) Greenstone, M., Gayer, T., (2007), Quasi-Experimental and Experimental Approaches to Environmental Economics, RFF Discussion Paper 07-22.

(L&L) Levitt, S., D., List, J., A. (2009), Field experiments in economics: The past, the present, and the future, European Economic Review 53, 1-18

Wikipedia:

In scientific inquiry, an ***experiment*** *(Latin: ex- periri, "to try out") is a method of investigating particular types of research questions or solving specific problems. An experiment is a cornerstone in the empirical approach to acquiring deeper knowledge about the world and is used in both natural sciences as well as in social sciences. An experiment is defined, in science, as a method of investigating less known fields, solving practical problems and supporting or negating theoretical assumptions.*

Brief History (L&L)

1) **Natural science experiments:** Among the first experimenters

    - **Galileo Galilei** in the 17th century: The falling bodies experiment (dropping objects from the leaning tower of Pisa in order to prove that all objects fall at the same rate, whatever their mass to disprove Aristotle's assertion that heavier bodies fall faster than light ones.),
    - **Sir Isaac Newton** showed that the white light is a mixture of colored lights and shattered down another Aristotle's theory that the white light is equal to purity,
    - **Pasteur** rejected the theory of spontaneous generation with an experiment – he showed that microorganisms grow in boiled nutrient broth when exposed to the air, but not when exposed to the carefully filtered air.

2) **"The dawn of field experimentation" in 1920s - 1930s:** Experiments were used to help to answer important economic questions. None of those studies involved human subjects. Via experimentation with agricultural plots Neyman and Fisher conceptualized RANDOMIZATION as a mean to achieve identification.

3) **Large-scale social experiments in mid 20th century** conducted by government agencies that involved individuals to evaluate employment programs, electricity pricing schemes, housing allowances … Experiments were used to test new programs later also reforms to existing programs – important influence on policy making.

    Social Experiment:

    Ferber and Hirsch (1982,p.7) '' a publicly funded study that incorporates a rigorous statistical design and whose experimental aspects are applied over a period of time to one or more *segments of a human population, with the aim of evaluating the aggregate economic and social effects of the experimental treatments.''*

Greenberg and Shroder (2004) define a social experiment as having at least the following four features: (i) random assignment, (ii) policy intervention, (iii) follow-up data collection, and (iv) evaluation.

4) **Growing popularity and expansion to diverse areas of interest over the past decade** economists have increasingly used the field experiments and controlled small(er)-scale experiments to explore economic phenomena.

Use of experimental methods in economics:

- education and training, public finance, industrial organization, labor and public economics, consumer behavior, game theory, development economics, **environmental economics** (see more in L&L and G&G)

Benefits of employing Experimental Methods

- A new drug is tested to make sure that it has the expected effect and at the same time that it is not outweighed by possible side-effects – to minimize potential cost on public health
- The effect of planned policy change can be tested at relatively low cost (compared to allocation of much larger resources to an inefficient program; e.g. training program for the unemployed, new pricing scheme for electricity,…)
- Explaining or predicting non-experimental outcomes (e.g. Barr and Serneels 2004: correlation of wage outcomes of employees with their behavior in a trust game experiment) – again, relevant policy/strategy implications at relatively low cost
- Testing theoretical predictions at relatively low cost (economic theory)
- Help to generate the data which are difficult to be obtained from "the field" (e.g. firms' or
- Estimation of a cost that the firm which produces pollution should internalize so that the (socially) more efficient outcome can be achieved **-> ENVIRONMENTAL ECONOMICS**

Environmental Economics -> Externalities (G&G)

- air or water pollution as a byproduct of the production of marketable good
- imposes health costs on inhabitants and/or costs on the down-the-river company not internalized by the firm which is responsible for producing the pollution
- government intervention might help to maximize net (social) benefits/welfare – require reliable estimates of the costs and benefits => ENVIRONMENTAL ECONOMICS
- **EE** addresses the inefficiencies resulting from production externalities –> experimental and quasi-experimental methods
- hinge upon proper design, implementation, appropriate approach to the data analysis

Possible Difficulties when employing Experimental Methods

L&L:

"The aim of the researcher is to estimate a causal effect of some action (a new government program, change in price,…), i.e. how outcomes differ when the action is taken vs. when it is not."

"The fundamental difficulty that arises is that either the action is taken or it is not—we never directly observe what would have happened in an alternative universe where a different action is taken. Thus, the construction of a control group becomes critical. Although we cannot observe what your outcome would have been had you not been treated, we can, for instance, observe outcomes for other similar individuals who were not treated."

G&G:
Causal Hypothesis

Illustrative Examples:
1) observational study analyzing the use of estrogen replacement therapy (ERT) to maintain the menopausal symptoms and their potential dangers such as higher incidence of heart disease
2) testing the impact of a new regulation which restricts pollution that can be produced by a company; what is the impact on health

- want to test the treatment effect (like e.g. receiving drug vs. placebo, exposure to high vs. low pollution)
- outcome may or may not respond to the treatment(=drug/high pollution) effect (heart disease/ other health problems) - > every individual has two potential outcomes but only one can be in fact observed
- to isolate the effect of treatment – all other factors need to be held constant (ideally, we would want to observe the outcome for the same individual in both treatments – with <u>and</u> without drug/ exposed to high <u>and</u> to low pollution – not possible -> Fundamental problem of Causal Inference)

➔ can observe the health outcome for treated individuals (with ERT/high pollution = TREATMENT GROUP) and for not-treated (no ERT/low pollution = CONTROL GROUP) -> average difference in health outcome treated vs. untreated

➔ PROBLEM -> SELECTION BIAS: our individuals might have some "special characteristics" that affect both, selection to treatment AND the outcome (women with healthier lifestyles/ people with lower income living in more polluted areas) -> the effect of special characteristics can be, in some situations validly assumed zero, in other situations it can be controlled for -> the researchers need to be aware of it to be able to make valid inferences!

Sources of the data (or, types of experimental approaches) (L&L, Harrison and List 2004)

- Naturally occurring data (identification assumptions) – selection into treatment is not up to researcher (e.g. ex-post examination of ERT, pollution regulation)
- Controlled data - selection into treatment is up to researcher (e.g. drug vs. placebo)
  - Field experiments -
    - Artefactual - departure from laboratory experiments; use "non-standard" subjects, or experimental participants from the market of interest, subjects understand that they are participating in an experiment
    - Framed - same as an artefactual field experiment but with field context
    - Natural - same as a framed field experiment but in the natural environment (where the subjects naturally undertake such tasks); the subjects do not know that they are participants in an experiment
  - Lab experiments (in laboratory setting, often using student subjects, randomization to identify the treatment effect)

A. Natural experiments = Quasi-experimental approaches (G&G)

SELECTION BIAS:

The researcher also compares the outcomes between the treatment and the control group. The difference is that the selection for the treatment is not up to the researcher – individuals are exposed (to treatment) by nature, politics, accident … -> NON-RANDOM ASSIGNMENT => possible source of SELECTION BIAS

➔ Still can make VALID inferences under the assumption that the assignment to the treatment is not related to any determinant of the outcome

- **Neg. Exmample 1:** Observational studies of ERT concluded no direct causality between ERT and heart disease. **Problem:** Maybe women with healthier life style were more likely to participate, take ERT and therefore per se less likely to have heart problems (reasons to believe that -> self-selection -> special characteristics that might affect the results of an observational study **Solution:** A randomized study -> Concluded that ERT substantially contributes to heart disease

- **Neg. Example 2:** similarly, in the pollution example if the housing prices are significantly lower in the affect area and therefore it is colonized by poorer people with less healthy lifestyles, less resources to spend on healthcare

- **Pos. Example:** If e.g. the government decides to enroll the unemployed in a special training program and selects randomly (or by some other rule, completely unrelated to their profession, abilities, education… anything that might affect their probability of success on the job market after the training) a half of the currently unemployed to receive the training (only a half for e.g. budgetary reasons, randomly to avoid e.g. accusations of discrimination) – their success after the training (if received) is not correlated with their selection for the group even though the assignment to treatment is not in the control of the researcher who will analyze the data

MEASUREMENT ERRORS:

- more sensitive people (health wise) might have migrated from the more polluted area – the time of their exposure is shorter than of individuals that stayed

- difficult to measure exact "exposure to pollution" (if interested in long term effects) over time (it might have varied)

- exposure to pollution might also vary location-vise (some people live closer to the source than other)

⇨ need to be very careful and employ proper techniques for data analysis, taking into consideration all possible biases and errors resulting from the nature of the experiment, concerned individuals and their selection for the experiment, their unobservable characteristics that might affect their "outcome"

Possible threats to the validity of the Q-E methods:

- Internal validity: treatment status can be related to the outcome for reasons other than treatment (selection bias)

- External validity: Heterogeneity of the treatment effect – possible problems with generalization of the result and application to other context. The estimated treatment effect might be different…
  - … from the overall population (perhaps the subjects are more sensitive to pollution)

- … across geographic settings
- … across institutional settings
- … across years

Example: If a government implements a program that improves the quality of air only in some regions, some individuals might change their location. If we estimate a short-run treatment effect before the re-location that would likely be different from the estimated long-run treatment after the re-location (large-scale experiment).

- Construct validity: the researcher must understand the experiment and the possible effects very well. In the pollution example, if the implemented program reduces the concentration of a specific air pollutant but at the same time has some other effect on health, positive or negative, the researcher is not aware of then he cannot separate the two (or more) effects by looking at the impact on health. Still can analyze the effect of the program on overall improvement in health. The researcher just has to be very careful in the specification of his research question and choose an appropriate methodology to get valid answer to that question.

What can we do to maximize the probability of the Q-E's validity

- To check the distribution of the observable covariates – if they are balanced than it may be reasonable to assume that also the unobservables are balanced in our sample
- Test whether the estimated effect is sensitive to a change in specification

➔ Still not a guarantee because the unobservables may still differ across the treatment and the control groups – therefore good understanding of the data (and their source) and economic intuition, reasoning or underlying model are very important and helpful!

B. Controlled Experimends (Field and Lab)

Use **RANDOMIZATION** to avoid the problem of **SELECTION BIAS** -> **Randomized experiments**

So what if there are problems with the data analysis that cannot be solved by a choice of proper approach to the data analysis? (e.g. we do not know what is it, what unobservable characteristic different than the exposure to treatment makes the women in our medical trial who take ERT less susceptible to heart disease -> therefore we cannot control for that factor).

➔ A classical experiment where subjects are randomly selected for treatment -> on average, individuals in treatment and in control group have (statistically) the same characteristics except of exposure to the treatments -> it is no longer the women themselves that decide whether to take the ERT or not, now it becomes to be under the control of the experimenter (a typical medical trial part of the patients, randomly selected, receives the real drug, the rest receives placebo) -> the **selection bias** disappears and the comparison of the outcomes in the two groups gives a credible estimate of the average effect of the treatment.

The use of randomized experiments in economics is growing rapidly.

RANDOMIZATION BIAS (L&L)

- some individuals might be reluctant to subject themselves to a random assignment =>

  experimental sample might differ from the population of interest because of randomization. For example, in medical trials, it is typically more difficult to persuade patients to participate in randomized than non-randomized studies; similarly, in social experiments the randomization might be difficult to implement (this could be a problem in both large and small-scale field experiments)
- participants in small-scale experiments might not be representative of individuals that would participate in a large-scale study; Heckman (1992), Heckman and Smith (1995), Manski (1995) (lab experiments, not natural field exp, when subjects are not aware of their participation)

SUBSTITUTION BIAS (L&L)

- subjects in the control group might seek available substitutes for treatment (large-scale experiments, NOT lab or framed experiments)

ATTRITION BIAS (L&L)

- within-subject design => some social experiments can be going on for several years during which subjects are surveyed – subjects might become tired of keeping detailed records, some might move,... (large-scale social experiments, not lab or framed experiments which are typically short-term)

GENERALIZATION OF THE RESULTS (L&L)

- even with proper estimation of the treatment effect, the generalizations of the results to other domains might prove difficult (lab experiment, framed experiments, NOT natural field experiments)
  - lab experiments: subjects are typically students, relatively small-scale
  - subjects are aware that they are monitored and recorded => psychological effect of being in the experiment, expecting the experimenter to expect specific result (see e.g. List 2006, or Benz and Meter 2008 for the difference in behavior when subjects are and are not aware they are participating in an experiment)

PUBLICATION BIAS (G&G)

- researchers more likely to submit, and journals are more likely to accept, for publication the studies that confirm the "expected results" (e.g. pollution is detrimental to health) – solution in leading medical journals, the researchers have to register their clinical trials, their study before knowing the results

REGULATORY BIAS (G&G)

- regulators put more weight on results that find a negative impact on health (to protect the public they require stronger evidence to support the " no-risk-to-health" than the "risk-to-health" results -> overestimated risk than reduces the chances to achieve the most efficient outcomes (risk-aversion is reflected in the willingness to pay and thereby affects the policy benefit calculations -> over-regulation -> E.g. pollutant A may be more risky than pollutant B, but if studies over-estimate the riskiness of B at the end the policy maker might not choose the most efficient allocation of resources to reduce the pollution.