Econometric Impact Evaluation and Economic Development

I used to think correlation implied causation.  
Then I took a statistics class. Now I don’t.  
Sounds like the class helped.  
Well, maybe.

Source: http://xkcd.com/552/

EIB E247, The Fletcher School  
Professor Jenny C. Aker  
Tufts University  
Fall 2010
Class Overview

• Q&A about course basics: Who am I, learning objectives, approach, pre-requisites, deliverables
• An Introduction to Econometric Impact Evaluation
Q&A about Course Basics

- Who am I?
- What are the objectives and approach of this course?
- What are my expectations? What are yours?
- What will you get out of this course?
Who am I?
Who am I?

• Professor Jenny C. Aker
• Email:  Jenny.Aker@tufts.edu or jennaker@hotmail.com
• Webpage: http://sites.tufts.edu/jennyaker/
• Office:  Cabot 603C
• Office Hours:  TBD each week and by appointment
• E-mailing
Who are you?

- Your name
- Your school/department
- Whether you are a first or second year
- Your econometrics background
- How certain are you that you will take this course? (Certain, fairly certain, not certain, I’m just here for the jokes)
- Why are you interested on econometric impact evaluation for development?
What is the objective of this course?

• The objective of this course is to provide students with a set of theoretical, econometric and practical skills to estimate the causal impact of one variable on another, with a particular focus on development programs.
Learning Outcomes

- Understand the value and practice of impact evaluation within the development community.
- Understand and apply a variety of econometric methods for estimating impact.
- Critically analyze impact evaluation research in economics and gauge the validity of the findings.
- Understand and apply evaluation design for development projects.
- Calculate the costs and benefits to different development interventions.
- Analyze existing data from a development project using impact evaluation techniques.
Course Approach

• Class lectures on technical topics
• Discussion of research papers, articles and blogs on impact evaluation
• In-class case studies
• Blackboard will have the readings, lectures and other resources (see “External Links” and “Calendar” for key dates)
• What about computers? Leave ‘em.
Resources


Pre-requisites

- Pre-requisites: Introductory econometrics at the level of EIB E213

- Econometrics may not be taken concurrently with this course, as lectures and assignments will assume a certain level of econometrics and STATA knowledge that will not be covered until later in EIB E213.
## Requirements

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four problem sets</td>
<td>35%</td>
</tr>
<tr>
<td>Three quizzes</td>
<td>25%</td>
</tr>
<tr>
<td><strong>Final project presentation and consultancy report</strong></td>
<td><strong>40%</strong></td>
</tr>
</tbody>
</table>
Group Project

• Groups of no more than 5 people
• Choose from among one of three development projects
• Clean the datasets*
• Analyze the data using one of the econometric techniques taught in class
• Present the results at a final colloquium (December 10th) and submit a consultancy report to the NGO
What will this course cover?

• I  Introduction to Impact evaluations
• II Basic Econometrics of Impact Evaluations
• III Randomized evaluations
• IV Regression discontinuity design
• V  Matching and Propensity Score
• VI Difference-in-differences
• VII Instrumental Variables
• VIII Cost-Benefit Analyses
• IX External Validity
Important Dates

• Partial class October 7th, make-up October 8th
• Potential no class October 21st, make-up November 8th
• No class November 18th, make-up November 22nd
• Final colloquium on Friday, December 10th (attendance required)
Class Overview

• What do we know? What don’t we know?
• What is impact evaluation and how is it different than “typical” evaluations in development?
• How do we determine causality?

Counterfactuals

• What are the characteristics of good counterfactuals?
• How do we choose counterfactuals?
• How can impact evaluations be used?
Do we know...
How to motivate this teacher...
To become this teacher?
How to transform these soils...
Into these soils?
How to ensure that this child...
Becomes this child?
Of course we know! We’re the « experts »

• These are complex and difficult questions
• We use experience and our judgment to determine the “best” approach
  o A scholarship for students
  o Demi-lunes for degraded soils
  o Cash for work to respond to a food crisis
  o But is there another subsidy, approach or package of interventions that could improve impact better and for a lower cost?
Choosing an Intervention is Complex...

We take a few « big » decisions during the project design phase, but other important decisions occur during the implementation.
And maybe not so simple.

- “After 2.3 trillion over 5 decades, why are the desperate needs of the world’s poor still so tragically unmet? Isn’t it finally time for an end to the impunity of foreign aid? — Bill Easterly, *The White Man’s Burden*
- “I have identified the specific investments that are needed [to end poverty]; found ways to plan and implement them; [and] shown that they can be affordable. — Jeffrey Sachs, *End of Poverty*
- “Rigorous impact evaluations of social development programs are relatively rare.” — Center for Global Development, *When Will we Ever Learn?*
Or is it?

“Success depends on knowing what works.”

Bill Gates
How can we choose the best intervention? Design, Monitoring and Evaluation

Source: CARE
Evaluation

• **Evaluation** is defined “the collection and analysis of information about the activities, characteristics and outcomes of programs” (Source: IRC *Evaluation Guidelines 2007*)
Evaluation

• Why do we do evaluations?
  o Accountability (Did we say what we said that we would do? Did we have a positive impact on people’s lives?)
  o Learning (What works? What doesn’t? Which is the most effective strategy)?

• What are characteristics of a good evaluation?
  o Answers an important question
  o In an unbiased and definitive way
  o But this requires a *theory of change*
The Theory of Change

<table>
<thead>
<tr>
<th>Needs</th>
<th>Logical Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor students living in the Busia District of Kenya have low incomes and low educational outcomes. They are unable to take books home to read and practice what they have learned in school.</td>
<td><strong>Inputs</strong></td>
</tr>
<tr>
<td><strong>NGO purchases textbooks and flipcharts for the schools</strong></td>
<td><strong>Books and flipcharts are delivered to the schools</strong></td>
</tr>
</tbody>
</table>
The Theory of Change

Source: Catholic Relief Services Propack
Different Types of Evaluation

- Needs assessments
- Process evaluation
- Impact Evaluation
- Cost-benefit analyses*
Needs Assessment

• Who is the target population?
  ○ All children? Those who live in the poorest areas? Those with the lowest test scores?

• What is the nature of the problem?
  ○ Low educational outcomes? Low enrolment? Low attendance?

• Who else is intervening in the area?

• What are potential interventions to address the problem?
  ○ Which services fit the environment?
Process Evaluation

• Has the program achieved its basic aims?

• In other words:
  o Are the services being delivered? Are they reaching the right population? Are the clients or beneficiaries satisfied?

• Process evaluation is often linked with monitoring ("monitoring and evaluation")
  o Monitoring is regular data collection on lower-level indicators to ensure that the project is “on the right track”
  o Evaluation is less frequent data collection on higher-level indicators
Impact Evaluation

• Has the program achieved its basic aims (outcomes)?

• Are these changes in outcomes due to the program? (causality or attribution)
  o How would individuals who participated in the project have fared in the absence of the project?
  o How would people who did not participate in the project have fared if they had participated in it?
  o This is the counterfactual or “missing data” problem
M&E versus Impact Evaluation

S&E
• Has the program been implemented in an efficient manner?
• Did the program target the right population?
• Is the project and its results on the right track?

Impact Evaluation
• What was the effect of the program on outcomes?
• To what extent did the well-being of beneficiaries improve due to the program?
• To what extent would the outcomes change if the project design were different?
• Is the program cost-effective?
M&E versus Impact Evaluation

• Were school scholarships delivered on time?  
  • M&E

• What are the trends in agricultural yields in Niger?  
  • M&E

• Do agriculture extension agents increase farmers’ adoption of new technologies?  
  • Impact evaluation
How do we usually do process evaluations?

• Think about a time when you participated in an evaluation of a project.

• What type of project was it?

• What indicator was the project evaluation trying to measure?

• How did you measure the project’s impact? In other words, from whom did you collect the data? When did you collect the data?
How do we do process evaluations?

- We often do “before-after” interventions
- In other words, we conduct a baseline study with beneficiaries before and compare the indicators before and after the project
- What are the problems with this approach?
Before-After Evaluations

• A USAID Title II program in Niger has the objective of increasing cowpea production in Niger. The project provides financial assistance to farmers to buy inputs (seeds, fertilizers, bags).
  o What is the theory of change?

• A process evaluation compares the cowpea harvest data before the project (2000) and after the project (2005). 2005 was one of the worst rainfall years in Niger on record for the past ten years. The cowpea production data are lower after the project than before.
  o Did the program fail?
Before-After Evaluations

- The program effect (financial assistance) cannot be distinguished from the rainfall effect.

How could we show that the program didn’t fail?
Before-After Evaluations

• The Minister of Rural Development in Niger has a project to improve livestock marketing between Niger and Nigeria. The objective of the project is to increase the prices received by herders.

• The project gives mobile phones to herders so that they can receive livestock prices by SMS from different markets.

• A before-after evaluation compares the prices received by the herders before and after the project. The indicators shows that prices are 500 CFA (about $1 USD) higher after the project.
Livestock Prices in Niger and Nigeria
Before-After Evaluations

- Did the program succeed?
- We don’t know
- How can we show that the project was successful in increasing herders’ prices more than the increase in livestock prices?
Isn’t it obvious?

- Investments in education lead to better educational outcomes, right? Visual aids are the answer!
- Why would they work? Why wouldn’t they work?
- Observational (retrospective) studies have compared schools with different levels of inputs and found that flip charts increased test scores by 20 percent.
Table 1
Retrospective estimates of effect of four flip charts in grades 6–8

Dependent variable: normalized 1998 test scores

<table>
<thead>
<tr>
<th>Specification</th>
<th>Mean(S.D.)</th>
<th>Level estimates</th>
<th>Diffs-in-diffs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Random effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>School × subject</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Schools</td>
<td>83</td>
<td>79</td>
<td>79</td>
</tr>
<tr>
<td>Pupils</td>
<td>5152</td>
<td>4998</td>
<td>4998</td>
</tr>
<tr>
<td>Grades included</td>
<td>6–8</td>
<td>6–8</td>
<td>6–8</td>
</tr>
<tr>
<td>Subjects included</td>
<td>Sc, Mat,</td>
<td>Sc, Mat,</td>
<td>Sc, Mat,</td>
</tr>
<tr>
<td></td>
<td>HS</td>
<td>HS</td>
<td>HS</td>
</tr>
</tbody>
</table>

Flip chart variable

| Number of charts in school    | 1.1 (2.4)  |
| (divided by four)            |            |
| Charts × flip-chart subject  | 0.192***   |
| (Science/ Agr., Math, HS–BE) |            |

0.080 0.065 0.064 0.041 0.057 0.056 0.049** 0.040* (0.021) (0.024)
Table 4
Prospective estimates of effect of flip charts—single subject multi-test regressions

<table>
<thead>
<tr>
<th>Subject</th>
<th>Past perf. Controls</th>
<th>Flip-chart school Coeff.</th>
<th>S.E.</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Flip-chart subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Science/Agriculture</td>
<td>No</td>
<td>0.0005</td>
<td>0.0752</td>
<td>20,446</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0.0007</td>
<td>0.0591</td>
<td>20,441</td>
</tr>
<tr>
<td>Math</td>
<td>No</td>
<td>-0.0201</td>
<td>0.0600</td>
<td>20,441</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>-0.0212</td>
<td>0.0486</td>
<td>20,434</td>
</tr>
<tr>
<td>Health Science/Business Ed. (HS–BE)</td>
<td>No</td>
<td>-0.0295</td>
<td>0.0728</td>
<td>20,434</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>-0.0276</td>
<td>0.0559</td>
<td>20,450</td>
</tr>
<tr>
<td>Geography/History/Civics/Religious Ed. (GHC)</td>
<td>No</td>
<td>0.0018</td>
<td>0.0714</td>
<td>20,450</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>-0.0012</td>
<td>0.0553</td>
<td>20,450</td>
</tr>
<tr>
<td><strong>Non-flip-chart subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>No</td>
<td>0.0038</td>
<td>0.0737</td>
<td>20,433</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>-0.0100</td>
<td>0.0576</td>
<td>20,448</td>
</tr>
<tr>
<td>KiSwahili</td>
<td>No</td>
<td>0.0110</td>
<td>0.0790</td>
<td>20,417</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0.0146</td>
<td>0.0737</td>
<td>20,417</td>
</tr>
<tr>
<td>Arts/Crafts/Music (ACM)</td>
<td>No</td>
<td>-0.0679</td>
<td>0.0758</td>
<td>20,417</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>-0.0723</td>
<td>0.0589</td>
<td>20,417</td>
</tr>
<tr>
<td><strong>Memo</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math and Science; grades 6 and 7 in 1998 only</td>
<td>No</td>
<td>0.0508</td>
<td>0.0828</td>
<td>13,836</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0.0534</td>
<td>0.0655</td>
<td>13,836</td>
</tr>
</tbody>
</table>

Regressions include school and school × year random effects and test fixed effects. Past performance controls are controls for the school-average performance on the July 1996 practice exam.
Causality versus Correlation

• **Causality / Causation** (X → Y): A change in X *causes* a change in Y
  – A scholarship (X) increases school attendance (Y)
  – A conditional cash transfer (X) reduces child malnutrition (Y)

• **Correlation**: X and Y move together and are somehow related to each other
  ○ Covariance (σ_{X,Y})
  ○ Correlation coefficient (ρ_{X,Y} = σ_{X,Y}/σ_Xσ_Y)

• *We want to show causality and rule out simple correlation*
Not all Correlations are Causal....

• Why might X and Y might be correlated?
  o causality       X ———> Y
  o reverse causality Y ———> X
  o simultaneity    X ———> Y and Y ———> X
  o omitted variables Z ———> X and Z ———> Y
  o spurious correlation (just ‘cause)
As the planet heats up, there are fewer pirates.

Global Average Temperature Vs. Number of Pirates

- 1820: Global Average Temperature is 13°C
- 1860: Global Average Temperature is 14°C
- 1880: Global Average Temperature is 14.5°C
- 1920: Global Average Temperature is 15°C
- 1940: Global Average Temperature is 15.5°C
- 1980: Global Average Temperature is 16°C
- 2000: Global Average Temperature is 16.5°C

The graph shows a positive correlation between global average temperature and the number of pirates over time.
Causality and Program Impact

• The program impact represents the part of the change that is caused by the program

• What is the measure of impact here?
  o B-A is the overall change (yellow)
  o B’-A is the trend
  o B-B’ is the program impact
How do we determine causality?

Source: xkcd.com
Causality and Program Impact

• We must use a counterfactual (comparison group, untreated group, control)

• A counterfactual is what would have happened in the absence of the program (or policy, intervention, etc)

• To measure the impact of the program, we must estimate the difference between the treated and control group...

• We want to compare the same individual with or without the program at the same moment in time

• Is this possible?
Counterfactual

Compare impact in participating and non-participating villages

Compare impact in participating and non-participating households within the same village
Causality and Program Impact

• To what extent does an education program improve test scores?

• What is the test score of a student in the program as compared with one outside of the program?

• The program impact is:

\[ \alpha = (Y | d=1) - (Y | d=0) \]

\[ \alpha = (\text{Score} | \text{Participating Student}) - (\text{Score} | \text{Non-Participating Student}) \]

What’s the potential problem with this comparison?
Criteria for a Good Counterfactual

- We want the treated person, household, village or association to have the same characteristics as the person, household, village or association
  - The only difference is the participation in the program
- Why? If there is no other reason for the differences between the two groups, if we see differences in the two groups, then it is due to an intervention
Criteria for a Good Counterfactual

• We want to determine if inputs (fertilizers) improve yields
• We could compare farmers who use the fertilizers (treated) and those who don’t (control). In doing so, we find that the treated farmers have higher yields
• Are these two farmers the same?
• Can we conclude that the use of fertilizers increases yields? What other factors could explain the difference?
False Counterfactuals

1. Before-after:
   - Same individual (village, association) before and after the treatment
   - Is this a good counterfactual?

2. Non-participating individuals
   - Those who chose not to participate
   - Those to whom the program wasn’t offered
   - Is the non-participant a good counterfactual?
Why? Selection Bias

- **Selection Bias**: The population chose to participate for specific reasons – often these reasons are correlated with the outcome of interest
- We can’t identify the impact of the program from other factors or reasons
How can we choose the counterfactual?

- **Experimental techniques**
  - Randomization or natural experiments

- **Quasi-experimental techniques**
  - Regression discontinuity design

- **Non-experimental approaches**
  - Matching/propensity score
  - Difference-in-differences
  - Instrumental variables
  - Interrupted time series
**The Vote 2002 Campaign (Arcenaux, Gerber and Green 2004)**

- The “Vote 2002” campaign was a campaign designed to increase voter turnout in 2002
- The campaign made phone calls to ~60,000 individuals
- But, only 35,000 individuals were reached
- Research question: Did the campaign have a positive impact on voter turnout?
- Five methods were used to estimate the impact
The Vote 2002 Campaign (Arcencaux, Gerber and Green 2004)

• The model is:

\[ V_{ist} = \alpha_{ist} + \beta phone_{ist} + X_{ist} \gamma + \theta_s + \varepsilon_{ist} \]

• \( V_{ist} \) is a binary variable for voting in 1998 or 2002
• \( phone_{ist} \) is a binary variable for whether the person received a call for the “Vote 2002” campaign
• \( X_{ist} \) is a vector of individual-level characteristics (age, sex, income, etc)
• \( \theta_s \) is a variable for state fixed effects
• How can this model be estimated?
• What is the interpretation of the coefficient \( \beta \)?
Estimated impact (coefficient on “phone”)

- Simple difference*
- Multiple regression*
- Multiple regression with panel data*
- Matching*
- Randomized Experiment
Impact Evaluation: Why should we do it?

- Interventions or policies seem obvious. **Not always.**
- We already have the evidence, right? **Not always rigorous evidence.**
- Tradeoffs between:
  - Scaling and Evaluation
  - Time and Evaluation
  - Can be a false trade-off.
- **Learning is important for action!**
Why don’t we do impact evaluations?

- Knowledge is a public good...
- Higher (perceived) cost and limited donor funding
- Pressing (real-world) needs for project implementation
- Technical knowledge of impact evaluations required
- Overly scientific approach
- Ethical concerns regarding randomization*
- Incentives to discourage impact evaluations (NGOs)
- Publication bias (academics)
Who can do impact evaluations?

- International organizations
- Governments
- NGOs
- Research institutes and universities
  - But these should ideally collaborate with one of the above. If not then it is often just research for research’s sake.
How can impact evaluations be used?

• Test a new approach
• Decide whether to continue, modify or end a program
  o Adjust the benefits of an existing program
• Compare costs and benefits of different types of interventions
• Inform policy
Deworming in Kenya (Miguel and Kremer 2004)

- An impact evaluation was carried out in Kenya on deworming medication.
- Deworming reduced the rate of absenteism by 25 percent and had a positive impact on untreated students (positive externality).
- Student attendance increased by .14 year for each child with a cost of US$3.50 for each school year - efficient investment.
The impact evaluation team launched the NGO « Deworm the World », which now reaches more than 20 million children in 27 countries.

The Kenyan government, in partnership with DTW, treated more than 3 million children since 2009 as a strategy to increase school attendance and improve child health.
But are impact evaluations always used this way? (Blattman 2008)

- Limited external validity
- Underreporting of negative and non-zero results
- Limited use to implementers (development practitioners)
- Too much focus on the black box (“did we reduce malnutrition?”) rather than the causal mechanisms (“why did approach X reduce malnutrition more than approach Y?”)
Evaluation 2.0 (Blattman 2008)

- Understand **whether and why** something works
- Use **experimental and non-experimental** approaches
- Ensure **upward (donor) and downward accountability**
- Provide timely results and feedback
Readings for Next Time

• Ravallion, 2008, “Evaluating Anti-Poverty Programs.”

• Angrist and Pischke, Mostly Harmless Econometrics, Chapters 1 and 2.

• Shadish, Cook and Campbell, pp. 1-18.