Applied Economics Dictionary

This document provides a brief overview of some terms common in applied micro-economic research. We begin by defining Applied Microeconomics before moving on to common concepts like empirical strategy and identifying assumption.

This document was created with the help of ClaudeAI. Any mistake is primarily the fault of the robots, but also of me for not checking carefully enough. Since the robots are busy, please direct any comments or report mistakes to claes.backman@gmail.com.

Table of Contents

Applied Microeconomics

Technical Definition:
Applied Microeconomics is the application of microeconomic theory and empirical methods to analyze real-world phenomena, markets, and behaviors. It combines economic theory, econometric methods, and data analysis to study cause-and-effect relationships and evaluate policies at a disaggregated level.

Plain Language Explanation:
Think of applied microeconomics as the "detective work" of economics - using data and economic theory to understand how individuals, firms, and markets behave in the real world, and how policies affect their behavior.

Key Components:

1. Research Areas:
- Labor Economics (wages, employment, education)
- Industrial Organization (firm behavior, market structure)
- Public Economics (taxation, public goods)
- Development Economics (poverty, growth at micro level)
- Environmental Economics (pollution, resource use)
- Health Economics (healthcare markets, health behaviors)

2. Methodological Approaches:
- Program Evaluation
- Quasi-experimental Methods
- Structural Estimation
- Reduced-form Analysis
- Survey Design and Analysis

Example Studies:

1. Labor Market Analysis:
- Question: "Do minimum wage increases reduce employment?"
- Methods:
  * Difference-in-differences
  * Border discontinuity designs
  * Synthetic control methods

2. Education Policy:
- Question: "What is the return to an additional year of schooling?"
- Methods:
  * Instrumental Variables (compulsory schooling laws)
  * Regression Discontinuity (school entry rules)
  * Twin studies

Key Features:

1. Empirical Focus:
- Heavy emphasis on causal identification
- Data-driven analysis
- Statistical and econometric methods
- Importance of research design

2. Policy Relevance:
- Informs policy decisions
- Evaluates program effectiveness
- Provides evidence-based recommendations
- Considers implementation challenges

3. Theoretical Foundation:
- Guided by economic theory
- Tests theoretical predictions
- Develops new theoretical insights
- Bridges theory and evidence

Common Tools:

1. Data Analysis:
- Large administrative datasets
- Survey data
- Experimental data
- Proprietary data sources

2. Statistical Software:
- Stata, R, Python
- Database management
- Data visualization
- Statistical testing

3. Research Methods:
- Regression analysis
- Instrumental variables
- Fixed effects models
- Machine learning methods

Current Trends:

1. Increasing Use of:
- Big Data
- Machine Learning
- Administrative Data
- Field Experiments

2. Focus on:
- Credible identification
- External validity
- Heterogeneous effects
- Mechanism exploration

Empirical strategy

Technical Definition: An empirical strategy is a systematic research plan that outlines how you will use data and statistical methods to identify and estimate causal relationships or test theoretical predictions. It includes your identification strategy, data collection approach, and econometric methods chosen to address potential estimation challenges.

Plain Language Explanation: Think of an empirical strategy as your research "game plan" - it's how you'll move from having a research question to getting credible answers. It requires you to think through:

- How you'll measure your variables of interest
- What might confound your analysis
- Which statistical techniques will help you overcome these challenges
- How you'll validate your findings

Example: Let's say you want to study the effect of minimum wage increases on employment. Your empirical strategy might be:

1. Use a difference-in-differences approach
2. Exploit variation in minimum wage laws across neighboring counties
3. Control for county-level characteristics
4. Use employment data from the Quarterly Census of Employment and Wages
5. Test parallel trends assumptions
6. Conduct robustness checks using alternative specifications

Key Considerations:

- Your strategy should address internal validity (causality) and external validity (generalizability)
- It should explicitly discuss potential endogeneity concerns
- It needs to be feasible given data availability
- It should include plans for robustness checks
- It should acknowledge and address potential limitations

Identification strategy

Identification strategy is closely related to but distinct from empirical strategy.

Technical Definition: An identification strategy is the logical argument and corresponding methodology that allows researchers to credibly estimate causal effects from observational data, focusing specifically on how to isolate the causal parameter of interest from confounding factors. It explains why your approach will yield estimates that can be interpreted as causal rather than merely correlational.

Plain Language Explanation: Think of an identification strategy as your answer to the question "Why should anyone believe that your analysis reveals causal effects rather than just correlations?" It's about finding a source of variation in your explanatory variable that is:

1. As good as randomly assigned
2. Relevant to your variable of interest
3. Only affects your outcome through the channel you're studying

Example: Suppose you want to study the effect of education on earnings. A simple OLS regression would be problematic due to ability bias (more able people tend to get more education AND earn more). Here's an identification strategy using instrumental variables:

1. Use changes in compulsory schooling laws as an instrument for years of education
2. Argue that:
    - Law changes are exogenous to individual characteristics
    - Laws affect education levels (relevance)
    - Laws only affect earnings through their effect on education (exclusion restriction)
3. Use 2SLS estimation to recover the causal effect

Key Considerations:

- Must explicitly state and defend key assumptions
- Should address potential threats to identification
- Often involves trading off internal and external validity
- May require multiple approaches for robustness
- Should be falsifiable (i.e., you can test some implications)

Common Identification Strategies:

1. Instrumental Variables
2. Regression Discontinuity
3. Difference-in-Differences
4. Matching Methods
5. Natural Experiments

Identifying assumption

Technical Definition: Identifying assumptions are the critical conditions that must hold true for your estimation strategy to recover causal effects. These are typically untestable assumptions that, when combined with your data and estimation method, allow you to interpret your estimates as measuring the causal relationship you're interested in.

Plain Language Explanation: Think of identifying assumptions as the "leap of faith" you're asking your audience to make - but one that should be justified by economic theory, institutional knowledge, and supporting empirical evidence. These assumptions fill the logical gap between what you can observe in your data and the causal effect you want to measure.

Example: Let's consider a difference-in-differences (DiD) analysis of a policy change:
Key Identifying Assumption: Parallel Trends

- Treatment and control groups would have followed parallel paths in the absence of treatment
- You can support this by:
    1. Showing parallel pre-trends graphically
    2. Running placebo tests on pre-treatment periods
    3. Conducting event studies
    4. Testing for balance in covariates

Specific Examples of Identifying Assumptions in Different Methods:

1. Instrumental Variables (IV):
    - Exclusion restriction: instrument affects outcome only through treatment
    - Relevance: instrument significantly affects treatment
    - Monotonicity: instrument affects all units in same direction
2. Regression Discontinuity (RD):
    - Continuity: all relevant factors except treatment vary smoothly around cutoff
    - No manipulation: units cannot precisely control their position relative to cutoff

Key Considerations:

- While identifying assumptions are untestable, you can often provide indirect evidence supporting them
- Stronger assumptions allow for stronger causal claims but are less likely to hold
- Multiple identification strategies with different assumptions can strengthen your case
- You should explicitly discuss what would violate your assumptions

Omitted Variable Bias

Technical Definition: Omitted Variable Bias is the bias that occurs in regression estimates when a relevant explanatory variable that is correlated with both the dependent variable and one or more included independent variables is excluded from the regression model. The bias arises because the included variables "pick up" some of the effect that should be attributed to the omitted variable.

Plain Language Explanation: Think of OVB as a "missing piece" problem. When you leave out an important variable that's related to both what you're studying (dependent variable) and your explanatory variables, your results will incorrectly attribute some of the missing variable's effect to your included variables, leading to biased estimates.

Example: Let's consider estimating the returns to education:
True Model: Wage = $\beta_0 + \beta_1$Education + β_2Ability + ε
Estimated Model (omitting ability): Wage = $\beta_0 + \beta_1$Education + ε

The Problem:

1. Ability affects both wages (↑) and education (↑)
2. By omitting ability:
    - Education coefficient is biased upward
    - It captures both direct effect of education AND indirect effect of ability
    - Formula for bias: bias = (Cov(Educ,Ability)/Var(Educ)) × β_2

Direction of Bias: The direction depends on:

1. Correlation between omitted and included variable
2. Effect of omitted variable on dependent variable

- If both are positive (or both negative) → Upward bias
- If they have opposite signs → Downward bias

Key Considerations:

- OVB is a key motivation for many identification strategies
- Solutions include:
    1. Finding proxy variables
    2. Using fixed effects
    3. Instrumental variables
    4. Natural experiments

Treatment Effects

Technical Definition: Treatment effects are parameters that measure the causal effect of an intervention (treatment) on an outcome of interest. They capture how outcomes change due to treatment compared to what would have happened without treatment.

Plain Language Explanation: Treatment effects tell us "what difference did the treatment make?" Different treatment effect parameters answer this question for different populations or under different conditions.

Key Types of Treatment Effects:

1. Average Treatment Effect (ATE):
- Definition: Expected effect of treatment for a randomly chosen unit from the population
- Formula: ATE = $E[Y(1) - Y(0)]$
- Example: Average effect of job training on all workers' wages, whether or not they would actually take the training

2. Average Treatment Effect on the Treated (ATT/ATET):
- Definition: Expected effect for those who actually received treatment
- Formula: ATT = $E[Y(1) - Y(0)|D=1]$
- Example: Effect of job training on wages for those who chose to take the training
- Often different from ATE due to selection into treatment

3. Local Average Treatment Effect (LATE):
- Definition: Average effect for "compliers" - units whose treatment status is affected by an instrument
- Context: Specific to instrumental variables estimation
- Example: Using college proximity as instrument for education, LATE measures returns to education for those who attend college only if they live near one

4. Intention to Treat (ITT):
- Definition: Effect of being offered treatment, regardless of whether treatment was actually received
- Usage: Common in RCTs with imperfect compliance
- Example: Effect of being offered a training program, even if some don't attend

Key Considerations:

1. Heterogeneity:
- Treatment effects often vary across units
- Different parameters might be relevant for different policy questions
2. Estimation Challenges:
- Fundamental problem of causal inference: can't observe both treated and untreated outcomes for same unit
- Selection bias can make some parameters harder to identify than others
3. External Validity:
- LATE might not generalize to other populations
- ATT might differ substantially from ATE if selection into treatment is important

4. Policy Relevance:

- ITT often relevant for policy evaluation
- ATT might be more relevant for program evaluation
- ATE crucial for universal policy considerations

Selection

Technical Definition:
Selection bias occurs when the sample or treatment group is not randomly selected from the population of interest, leading to systematic differences between groups that affect the outcome. This results in biased estimates because the selection process is correlated with the outcome of interest.

Plain Language Explanation:
Selection bias happens when the way people/units end up in your sample or choose to participate in a program is systematically related to the outcomes you're studying. It's like trying to estimate average height by only measuring basketball players - you'd get a biased estimate.

Types of Selection Bias:

1. Self-Selection Bias:
- When individuals choose whether to participate
- Example: Studying returns to MBA education
 * People who choose to get MBAs might be more ambitious/capable
 * Higher post-MBA salaries might reflect both MBA effect AND pre-existing characteristics

2. Sample Selection Bias:
- When data is only observable for a non-random subset
- Classic Example: Women's wages (Heckman)
 * Only observe wages for women who choose to work
 * Working women might have systematically different characteristics

3. Attrition Bias:
- When subjects drop out of study non-randomly
- Example: Long-term education studies
 * If lower-performing students more likely to drop out
 * Results will overstate educational effectiveness

Key Considerations:
1. Detection:
- Compare sample/participant characteristics to population
- Look for systematic patterns in missing data
- Test for selection on observables

2. Documentation:
- Be explicit about potential selection mechanisms
- Discuss direction of likely bias
- Consider multiple correction methods

3. Limitations:
- Selection on unobservables particularly challenging

- Correction methods often require strong assumptions
- External validity concerns remain

Reverse Causality

Technical Definition:
Reverse causality (also called reverse causation or simultaneity) occurs when the presumed effect might actually cause the presumed cause, or when two variables simultaneously cause each other. This creates an endogeneity problem that biases standard regression estimates.

Plain Language Explanation:
While you might think A causes B, it could be that B causes A, or they might cause each other simultaneously. This makes it difficult to isolate the true causal effect in either direction.

Example:
Consider studying the relationship between crime and police officers in a city:

1. Hypothesized Relationship:
- More police officers → Less crime

2. Reverse Causality Problem:
- Cities with more crime → Hire more police officers
- Naive regression might show positive correlation between police and crime
- This correlation reflects reverse causation, not causal effect

Common Examples:
1. Education and Income:
- Does education increase income?
- Or do wealthier families invest more in education?
- Likely both: simultaneous causation

2. Health and Wealth:
- Does wealth improve health outcomes?
- Or do health problems reduce earning capacity?
- Probably bidirectional

Solutions:

1. Instrumental Variables:
- Find variable that affects X but not Y directly
- Example for police-crime: Use timing of terror alerts as instrument for police deployment

2. Timing/Lags:
- Use lagged values of independent variables
- Relies on assumption that future values don't affect past values
- Example: Use last year's police levels to predict this year's crime

3. Natural Experiments:
- Find exogenous variation in X
- Example: Use weather shocks to study effect of agricultural output on conflict

4. Simultaneous Equations:
- Model both relationships explicitly
- Requires identifying assumptions
- Often uses structural models

Key Considerations:

1. Identification:
- Need to justify why causation runs one way
- Look for timing differences
- Consider institutional details

2. Testing:
- Granger causality tests
- Specification tests
- Falsification tests

3. Common Pitfalls:
- Timing might not solve problem if variables are forward-looking
- Hard to find good instruments
- Simultaneous equations need strong assumptions

Natural Experiments

Technical Definition:
Natural experiments (also called quasi-experiments) are situations where variation in the treatment variable occurs due to natural circumstances, policy changes, or institutional rules that are plausibly exogenous to the outcome of interest. These situations approximate random assignment, allowing researchers to estimate causal effects.

Plain Language Explanation:
Think of natural experiments as scenarios where "nature" (or policy, institutions, etc.) creates an experimental-like setting for us. Instead of researchers randomly assigning treatment, some external force creates variation that we can use to study causal effects.

Types of Natural Experiments:
1. Policy Changes:
- Example: Education effects using compulsory schooling laws
  * Different states change laws at different times
  * Creates variation in education levels
  * Assumes timing of law changes is exogenous

2. Geographic Discontinuities:
- Example: Minimum wage effects at state borders
  * Compare neighboring counties across state lines
  * Similar counties, different policies
  * Assumes border location is arbitrary

3. Administrative Rules:
- Example: Class size effects using maximum class size rules
  * Schools must create new class when enrollment exceeds threshold
  * Creates discontinuous variation in class size
  * Combines natural experiment with RD design

4. Natural Events:
- Example: Weather shocks on economic outcomes
  * Random variation in rainfall/temperature
  * Can affect agricultural productivity
  * Assumes weather is exogenous to outcomes

Key Elements of Good Natural Experiments:
1. As-Good-As-Random Assignment:
- Treatment variation should be plausibly exogenous
- Units can't select into treatment based on potential outcomes
- Pre-treatment characteristics should be balanced

2. Relevance:

- Treatment variation should meaningfully affect variable of interest
- Effect size should be large enough to detect
- First-stage relationship should be strong

3. Exclusion Restriction:
- Treatment variation should only affect outcome through proposed channel
- No direct effect on outcome
- No other channels of influence

Validation Strategies:

1. Balance Tests:
- Compare pre-treatment characteristics
- Should be similar across treatment/control
- Test for systematic differences

2. Placebo Tests:
- Test effect on unrelated outcomes
- Examine pre-treatment trends
- False experiments in different time periods

3. Robustness Checks:
- Different specifications
- Various control groups
- Alternative treatment definitions

Common Pitfalls:

1. External Validity:
- Effects might be local to specific context
- May not generalize to other settings
- Need to consider policy relevance

2. Anticipation Effects:
- Units might respond before actual change
- Can bias estimates
- Need to consider announcement effects

3. Concurrent Changes:
- Other policies might change simultaneously
- Need to account for multiple treatments
- Could confound causal interpretation