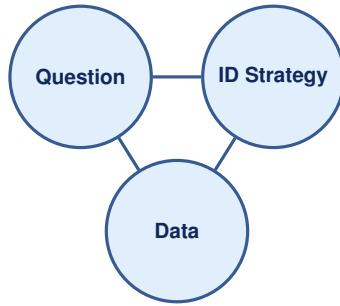


# Choosing Your Research Topic

14.33 Research & Communication in Economics | Theodore Caputi

## The Matching Game

Applied micro projects are a matching game. You need three things that fit together—a **question**, an **identification strategy**, and **data**—and they all have to line up. Think of it as a three-legged stool: if any leg is missing, the whole thing falls over.



**Don't panic about novelty.** Most economics projects take 5+ years and make only a marginal contribution. You don't need to revolutionize the field. You need an **interesting, feasible project** you can finish by the end of the semester. **Feasibility beats novelty.**

## Leg 1: The Question

For this class, the question is the easiest leg. Any “how does X affect Y?” question is fair game, as long as X and/or Y concern **human or social behavior**. That’s a broad mandate—health, education, crime, labor, the environment, technology, and much more are all fair game.

Your question needs to be:

- **Causal.** You’re asking whether some policy or event *changed* behavior. “Does X affect Y?” is the right framing—not “are X and Y correlated?”
- **Interesting to you.** You’ll spend weeks on this. Pick something you genuinely care about.

That’s really it. You don’t need a question that has never been asked. You need a clear causal question about human behavior that lines up with a feasible strategy and available data.

**Example questions:** Does raising the minimum wage reduce employment? Do texting bans reduce traffic fatalities? Does Medicaid expansion improve health outcomes? Does recreational marijuana legalization affect alcohol consumption? Do class size reductions improve test scores?

## Leg 2: The Identification Strategy

There are many approaches to causal inference—difference-in-differences (DD), instrumental variables, regression discontinuity, and others. I’ll use DD as my running example below, but you should feel free to use whatever strategy best fits your question and data. DD compares how an outcome changes over time for a

group affected by a policy (the *treatment group*) versus a group that wasn't (the *control group*). The cleanest DDs involve treatments with these three features:

### 2.1 A Discrete Policy or Event

The cleanest DD treatments are things that happened **all at once and lasted**—a clear before-and-after:

- A state or county enacts a new law on a specific date
- A federal program launches and changes behavior from that point on
- A major event (natural disaster, court ruling) that created a sharp break

**Bad example:** “The rise of social media.” That’s gradual, not discrete. You can’t draw a clean line between before and after.

### 2.2 The Treatment Affected Some Groups but Not Others

Your treatment needs to have affected some groups but not others—that’s what gives you your treatment and control groups. The easiest way to think about this is **geographically**—Medicaid expansion affected people in states that expanded Medicaid and didn’t affect people in states that didn’t. But you can also use other dimensions: maybe a policy only affects non-citizens and doesn’t affect citizens.

**What about spillovers?** A spillover is when the policy also affects the control group. You’ll never find anything with zero spillovers—Medicaid expansion could feasibly affect neighboring states. But you want to **minimize** spillovers. The cleaner the separation, the better.

### 2.3 The Treatment Is Exogenous

You want a treatment that is **exogenous**—one that originated outside “the system,” where the affected people didn’t play much of a role causing it. A useful thought experiment: think about a person affected by the treatment and ask how significant a role *they* played in causing it. The less of a role, the more exogenous the treatment is.

**Good example:** Most state policies are reasonably exogenous. An individual in Massachusetts didn’t single-handedly cause Medicaid expansion. Of course, nothing is 100% exogenous—state policies are adopted because residents want them—but an individual person has very little control over state policy.

**Bad example:** Do high SAT scores cause people to go to college, or do college-bound people just study harder for the SAT? Here the “treated” person played a direct role in their own treatment—the direction of causality is tangled. That’s *endogenous*.

## Leg 3: Data

---

For a DD, you need data on the outcome **before and after** the treatment in **both** the treatment and control groups. This sounds obvious, but it trips people up constantly.

### 3.1 Make Sure Pre-Treatment Data Exists

Before you fall in love with a question, verify that data exists for the period *before* the policy was implemented.

**Bad example:** Uber/Lyft legalization and ride-hailing trip counts. If your data comes from Uber or Lyft, you’ll only have data after they were legalized—so you have no pre-treatment data for the treated group, and no data at all for places that never legalized. No pre-treatment data means no DD.

### 3.2 Make Sure You Have Enough Post-Treatment Data

A policy that happened recently is usually too recent—publicly available data always lags behind real time.

**The 8–15 Year Rule.** Look at policies from roughly **8–15 years ago** rather than the last couple of years.

You have a much better chance of finding good post-treatment data for a 2012 policy than a 2024 one. I know many of you want to study current debates, but older policies are far more feasible for this class.

**How to find policy ideas:**

- MIT Libraries has access to historic newspaper archives—look at what policies were debated or enacted 8–15 years ago.
- Academic papers review major policy changes in their literature reviews—skim a few in a topic you like.

## Putting It All Together

---

**My recommendation: start with identification.** Find a policy or event that created plausibly exogenous variation. Then work outward from there.

### Turning a Treatment into a Project

Once you've found a policy or event that interests you, ask yourself three questions:

1. **Can I find data on the treatment?** You need to know when and where each unit (state, country, school district, etc.) was affected. Check academic papers in the area, NCSL, or other policy databases. If no one has compiled the details, you may have to do it yourself—that's a lot of work for a semester project.
2. **What outcome might this treatment affect?** Think about what the policy was designed to change, or what it might unintentionally affect. Texting bans → traffic fatalities. Minimum wage → employment. Marijuana legalization → drug use, crime, traffic accidents, etc.
3. **Can I find data on the outcome?** The outcome must exist in a dataset that covers the relevant units and time periods for your identification strategy. FARS for traffic deaths, BRFSS for health behaviors, CPS for employment, etc.

If you can answer “yes” to all three, you have a viable project.

### The Checklist

1. **Start with a topic that interests you.** Browse newspapers, skim papers, think about policies or events that have affected different groups differently.
2. **Check identification.** Do you have a credible strategy for estimating a causal effect? Is there a discrete event or policy that created variation you can exploit? Is the variation reasonably exogenous?
3. **Check data.** Does data exist for the comparison you need? Do you have enough observations in the right places and time periods?
4. **If all three line up, you have a viable project.** If not, iterate—tweak the question, try a different treatment, or look for different data.

The best project is one you find interesting, can identify causally, and can complete with available data.

**Feasibility beats novelty.**