Causal Inference

Intro to Controlling for Observables

Brad Hackinen

Agenda:

- Intro to Controlling for Observables
- A survey of adjustment methods

Recap: Measuring Causal Effects with Randomized Experiments

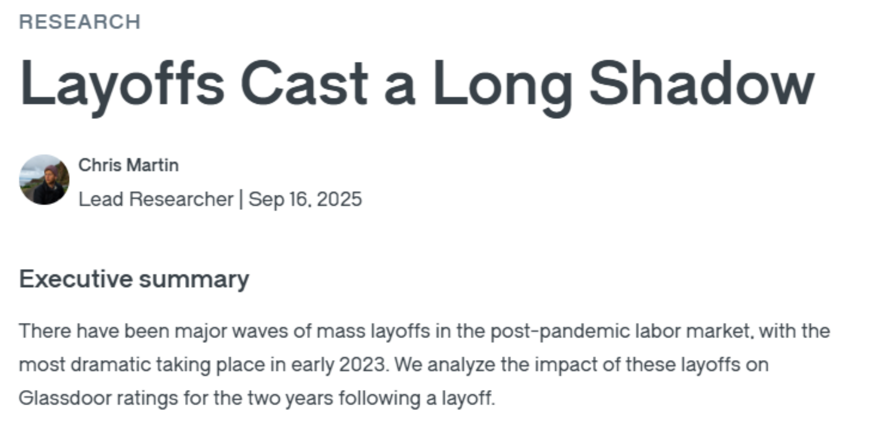
Q: Do randomized experiments solve the *fundamental problem of causal inference*?

Why Randomized Experiments Work

Randomized experiments are the “gold standard” for causal inference.

- When treatment is assigned independently of potential outcomes, we can estimate average counterfactual outcomes with the average observed outcomes of the appropriate group. This makes estimating average treatment effects very simple.
- In general, “treatment is assigned independently of potential outcomes” is an *assumption* we use to derive our estimator.
- However, deliberate randomization of treatment can make this a very safe assumption.

On the other hand. . . Could this study be run as a randomized experiment?



RESEARCH

Layoffs Cast a Long Shadow

Chris Martin
Lead Researcher | Sep 16, 2025

Executive summary

There have been major waves of mass layoffs in the post-pandemic labor market, with the most dramatic taking place in early 2023. We analyze the impact of these layoffs on Glassdoor ratings for the two years following a layoff.

Intro to Controlling for Observables

Recap

- Randomized treatment allows us to estimate causal effects by substituting observed outcomes in one group for the unobserved counterfactuals in another
- But, randomized experiments can be too costly or unethical to use in many real-world situations.

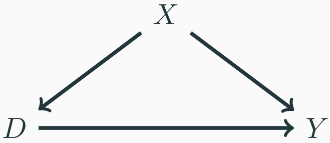
Oberservational Studies

- An *Observational Study* is a one where treatment is not assigned by the researcher, but is instead determined by some other process.
- When we try to measure causal effects in observational studies, we need to worry about *confounders*

Review: We call variable X a ***confounder*** if it is an ancestor of both D and Y.

Confounders can generate a statistical relationship between D and Y that is observationally equivalent to a direct effect $D \to Y$.

Example:



Real World Observational Data

Let's look at a real-world case study.

`NSWCaseStudy_blank.md`

Step 1: Understand the data

- Where does the data come from?
- What do the variables mean?
- Compute the difference-in-means estimate of the ATT
- How does it compare to the experimental estimate?

Work in small groups

`NSWCaseStudy_blank.md`

Step 2: Data exploration

1. "Explore" the data
   - How are the variables distributed? Do you see any interesting patterns?
2. Look for evidence of confounders
   - *Hint: What confounders might exist in this setting? Can any of them be observed? What patterns would they be likely to create?*
3. **Prepare an interesting plot to show**
   - Make a plot that illustrates a potential confounder in the data

Feel free to use AI tools to help you make plots, but make sure you understand what they show!

Controlling for Observables

- The goal of “controlling for observables” is to to make the data look *as if* we had had run an experiment with randomized treatment, even when we did not.
- If successful, controlling for observables allows us to estimate average treatment effects with observational data.
- The challenge is that we cannot observe everything about each unit in the data. Instead we only have access to a limited number of descriptive variables, which we call ***covariates*** or ***observables***.

“Controlling” vs “Adjusting”

Controlling is another area of Causal Inference where people use different language to describe the same thing.

The following statements are all equivalent:

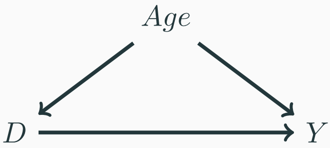
- *“We controlled for X,Y,Z”*
- *“Our control variables are X,Y, and Z”*
- *“We adjusted for X,Y and Z”*
- *“Our adjustment set is {X,Y,Z}”*

Which terminology you use is mostly a matter of taste.

A simple example

Suppose we have some data on a medical treatment where, on average:

- The outcomes are worse for older patients
- Older people are more likely to be treated than younger people



Thus, age is a confounder because it is a causal ancestor of both treatment and outcome.

Estimating the ATT

Suppose we want to estimate the average treatment effect on the treated (ATT) with the difference-in-means estimator:

The estimate will be:

$$\Delta Y = \overline{Y}_T(1) - \overline{Y}_U(0)$$

However, the true value of the ATT is:

$$ATT = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]$$

The *bias* of the estimate is the difference between the expected value of the estimate and the true value:

$$
\begin{aligned}
Bias &= E[\Delta Y] - ATT \\
&= E[\overline{Y}_T(1) - \overline{Y}_{NT}(0)] - (E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]) \\
&= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] \\
&\qquad - (E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]) \\
&= E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]
\end{aligned}
$$

Implication: The difference-in-means estimator will be biased if the observed outcomes for untreated units are not a good proxy for the counterfactual outcomes of treated units.

Confounding Bias

Confounders create correlations between treatment and outcome that are not due to a causal effect.

Another way to say this is that the potential outcomes in the treated and untreated groups are systematically different.

This means that the difference-in-means estimator will generally be biased when there are confounders.

Note: In economics and some social sciences, this is often called "*selection bias*".

Example: Age as a Confounder

Let's go back to our medical treatment example with age as a confounder:

- Older patients are more likely to be treated and also have worse outcomes on average, regardless of treatment.
- So on average, the untreated outcomes for treated patients will be worse than the untreated outcomes for untreated patients:

$$E[Y_i(0)|D_i = 1] < E[Y_i(0)|D_i = 0]$$

- This means that the bias will be negative and the difference-in-means estimator will *underestimate* the treatment effect.

Constructing a better Counterfactual

Suppose we focus on a specific treated unit, i, with age x_i and we want to estimate its counterfactual outcome, $Y_i(0)$.

Some options:

1. Find a single control unit j in the untreated group with the same age $(x_j = x_i)$ and use its outcome as an estimate of the counterfactual (Matching)
2. Use the average outcome of all control units with the same age x_i as an estimate of the counterfactual (Post-stratification/Weighting)
3. Fit a statistical/machine learning model to predict the outcome for untreated units as a function of age, and then use the model to predict the counterfactual outcome for unit i (G-computation/Regression)

Prediction as Conditioning as Controlling

When we use a confounding variable to help predict counterfactual outcomes, we are “controlling” for that variable.

More formally, what these procedures are doing is *conditioning* on the confounder when computing differences in expected outcomes.

Under the right conditions, this will give us an unbiased estimate of the true causal effect.

Formal Requirements

Mathematically, controlling for observables will give us an unbiased estimate of a causal effect if three assumptions hold:

1. Treatment is *conditionally independent* of the potential outcomes given X
2. $0 < P(D_i|X) < 1$ ("*Positivity*")
3. Stable Unit Treatment Value Assumption (SUTVA)

Conditional Independence Assumption (CIA)

The *Conditional Independence Assumption (CIA)* states that treatment is independent of the potential outcomes after conditioning on some set of observable covariates X.

In essence, CIA is a way of saying that treatment is effectively random after we control for certain observable covariates.

Selection on Observables

CIA will always hold if all confounders are *observable* (this is sometimes referred to as "*selection on observables*").

The problem is that we often need to control for *unobservable* variables.

Whether the CIA holds in practice will depend on the specific causal effect and available data.

Comment: *CIA is not testable from the data alone*. We will have to use our domain knowledge to judge whether it is reasonable in a particular context.

Positivity

The *Positivity* assumption states that for each value of X there is some non-zero probability of observing both treated and untreated units:

$$0 < P(D_i|X) < 1$$

Basically, to predict counterfactual outcomes for a given treated unit, we need to observe some untreated units with the same (or at least similar) characteristics. Otherwise we are really just guessing.

Common Support

In practice, the positivity requirement often does not hold for *all* values of X in our data. In these situations, we can still estimate causal effects for certain values of X where positivity holds.

The set of values for X where we observe both treated and untreated units is called the *common support*.

When CIA, Positivity, and SUTVA hold with respect to covariates X, we can construct a similar argument to when treatment is fully independent of the potential outcomes:

If D_i is independent of $Y_i(0)$ and $Y_i(1)$ conditional on X, then:

$$E[Y_i(0)|D_i = 1, X_i = X] = E[Y_i(0)|D_i = 0, X_i = X]$$
$$E[Y_i(1)|D_i = 0, X_i = X] = E[Y_i(1)|D_i = 1, X_i = X]$$

Which allows us to estimate the expected counterfactual outcomes from the observed outcomes by conditioning on X.

Approaches to Controlling for Observables

There are many of approaches to controlling for observables, including:

- Matching Estimators
- Weighting Estimators
- Regression
- Propensity Score Methods
- Causal Machine Learning
- Structural Models

These techniques are more similar than they sound: They can all be interpreted as ways of conditioning on control variables to create better predictions for counterfactual outcomes.

Back to our NSW data

`NSWCaseStudy_blank.md`

Goal:

- Use observable covariates to predict better counterfactuals and estimate the ATT

A Survey of Adjustment Methods

Five Big Ideas in Controlling for Observables

1. Matching
2. Weighting
3. Imputation
4. Regression
5. Structural Modeling

Matching

Matching methods attempt to estimate counterfactual outcomes by finding similar units in the data that received a different treatment.

- **Exact Matching**: Find one or more units in the other treatment group with exactly the same covariate values.
- **Approximate Matching**: Find one or more units in the other treatment group with “similar” covariate values (there are many different algorithms with different definitions of “similar”).

Common Support

The region of the covariate space where we can observe both treated and untreated units is called the ***common support***.

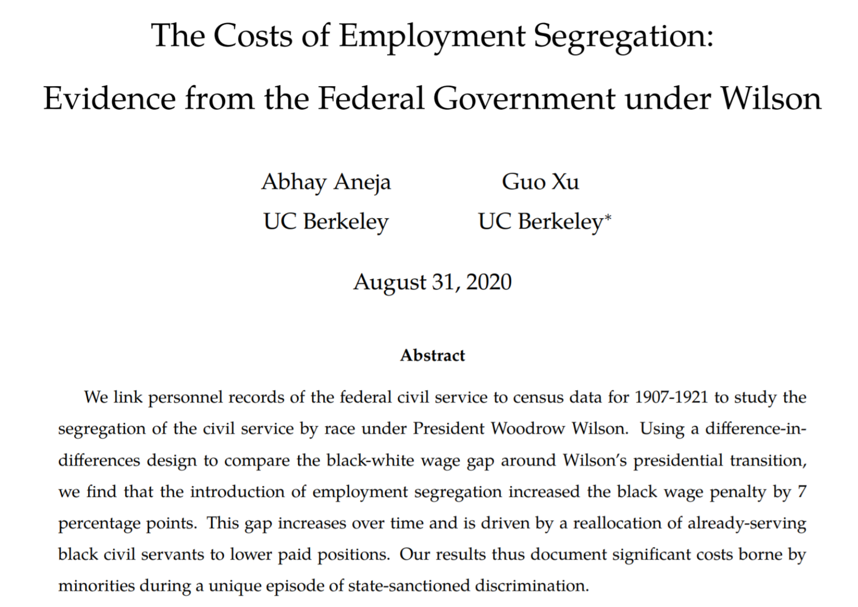
Matching estimators can only estimate treatment effects for units that lie in the common support.

Selecting an Estimand

Matching estimators can be used to estimate a variety of average treatment effects:

- To estimate the ATT, find one matched control for each treated unit
- To estimate the ATU, find one matched control for each untreated unit
- To estimate the ATE, find one match from the other treatment group for every observed unit

An example of causal inference with exact matching



The Costs of Employment Segregation:
Evidence from the Federal Government under Wilson

Abhay Aneja — UC Berkeley

Guo Xu — UC Berkeley*

August 31, 2020

Abstract

We link personnel records of the federal civil service to census data for 1907-1921 to study the segregation of the civil service by race under President Woodrow Wilson. Using a difference-in-differences design to compare the black-white wage gap around Wilson's presidential transition, we find that the introduction of employment segregation increased the black wage penalty by 7 percentage points. This gap increases over time and is driven by a reallocation of already-serving black civil servants to lower paid positions. Our results thus document significant costs borne by minorities during a unique episode of state-sanctioned discrimination.

This table shows differences in the distribution of wages between black and other civil servants in the raw and matched data:

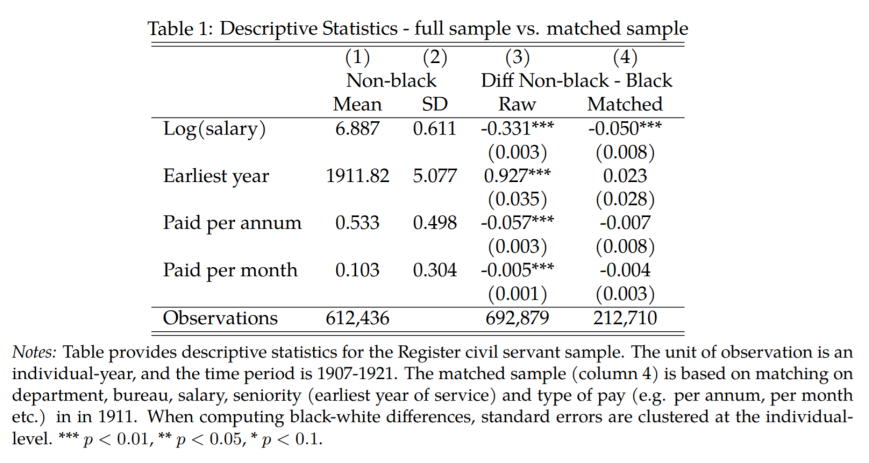
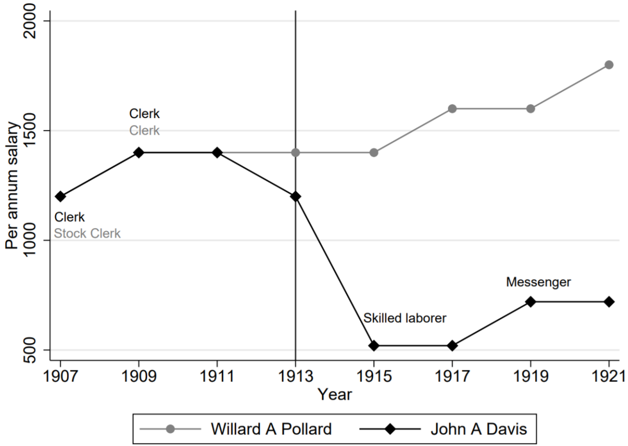


Table 1: Descriptive Statistics - full sample vs. matched sample

	(1)	(2)	(3)	(4)
	Non-black		Diff Non-black - Black	
	Mean	SD	Raw	Matched
Log(salary)	6.887	0.611	-0.331***	-0.050***
			(0.003)	(0.008)
Earliest year	1911.82	5.077	0.927***	0.023
			(0.035)	(0.028)
Paid per annum	0.533	0.498	-0.057***	-0.007
			(0.003)	(0.008)
Paid per month	0.103	0.304	-0.005***	-0.004
			(0.001)	(0.003)
Observations	612,436		692,879	212,710

Notes: Table provides descriptive statistics for the Register civil servant sample. The unit of observation is an individual-year, and the time period is 1907-1921. The matched sample (column 4) is based on matching on department, bureau, salary, seniority (earliest year of service) and type of pay (e.g. per annum, per month etc.) in in 1911. When computing black-white differences, standard errors are clustered at the individual-level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure 2: Example of matching strategy – John A Davis vs. Willard A Pollard



Notes: Figure illustrates the matching strategy for the case of John A Davis (black civil servant) and Willard**jy** A Pollard (white civil servant). The career progression shown for both civil servants is based on actual data on salary and occupational position from the federal Register. The solid line marks the year of Wilson's inauguration and implementation of the segregation order.

Look how the differences between the two groups are mostly eliminated by matching:

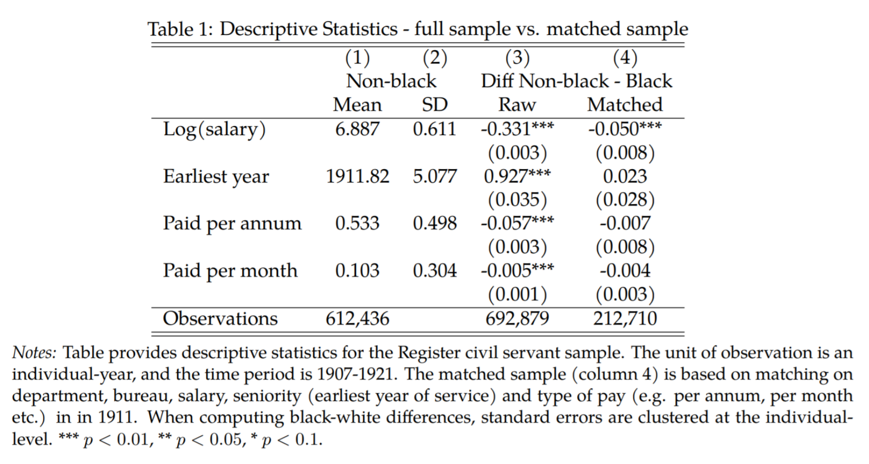


Table 1: Descriptive Statistics - full sample vs. matched sample

	(1)	(2)	(3)	(4)
	Non-black		Diff Non-black - Black	
	Mean	SD	Raw	Matched
Log(salary)	6.887	0.611	-0.331***	-0.050***
			(0.003)	(0.008)
Earliest year	1911.82	5.077	0.927***	0.023
			(0.035)	(0.028)
Paid per annum	0.533	0.498	-0.057***	-0.007
			(0.003)	(0.008)
Paid per month	0.103	0.304	-0.005***	-0.004
			(0.001)	(0.003)
Observations	612,436		692,879	212,710

Notes: Table provides descriptive statistics for the Register civil servant sample. The unit of observation is an individual-year, and the time period is 1907-1921. The matched sample (column 4) is based on matching on department, bureau, salary, seniority (earliest year of service) and type of pay (e.g. per annum, per month etc.) in in 1911. When computing black-white differences, standard errors are clustered at the individual-level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Tradeoffs in Matching

- Exact matching is often not feasible because only a tiny fraction of the data can be matched.
- Approximate matching allows us to find matches for a larger fraction of our sample observations.

. . . Unfortunately, as soon as we relax the constraint that observations need to be matched exactly, we also introduce the potential for our estimator to be *biased*. After all, we are only *approximately* conditioning on the matching variables.

Bias-Variance Trade-off

Matching estimators force us to make fundamental trade-off between *bias* and *variance*.

In general, choosing more restrictive matches leads to:

- Better matched controls within sample → reduced estimator bias
- Smaller sample → higher estimator variance, reduced external validity

Internal/External Validity Trade-off

Observations that can be matched are often systematically different than those that are hard to match. If the treatment effects are different between the two groups, using more restrictive matching will:

- Reduce bias when estimating the treatment effect for matched observations (better internal validity)
- Potentially *increase* bias if we apply the estimate to the unmatched observations (poorer external validity)

Weighting Estimators

Weighting estimators attempt to estimate counterfactual outcomes by re-weighting the observed outcomes of units in the other treatment group so that the weighted distribution of covariates matches the distribution of covariates in the treatment group of interest.

Weighting Example

Suppose we want to estimate the ATT in our example of age-based confounding. We can:

1. Estimate the mean observed outcome for treated units using and unweighted average:

$$\overline{Y}_T = \frac{1}{N_T}\sum_{i \in T} Y_i$$

2. Estimate the mean observed outcome for untreated units using a weighted average:

$$\overline{Y}_U^W = \frac{1}{\sum_{i \in U} w_i}\sum_{i \in U} w_i Y_i$$

where the weights w_i are chosen so that the weighted distribution of age in the untreated group matches the distribution of age in the treated group.

Weighting and the Common Support

Note that weighting estimators can only estimate treatment effects for units that lie in the common support.

For example, if there are treated individuals who are 70 years old, but no control individuals who are 70 years old, there is no weight we can choose to make the control group have the same fraction of (weighted) individuals who are 70.

Matching as Weighting

Matching can be interpreted as kind of weighting estimator where:

- Each observation is weighted according to the number of times it appears in a matched pair
- Observations outside of the common support are given zero weight

Choosing the weights

We choose the weights to mimic the distribution of observed covariates in the population we want to estimate the treatment effect for:

- If we choose weights to match the distribution of covariates in the treated group, we can estimate the ATT
- If we choose weights to match the distribution of covariates in the untreated group, we can estimate the ATU
- If we choose weights to match the distribution of covariates in the overall population, we can estimate the ATE

Imputation Methods

Imputation methods attempt to estimate missing counterfactual outcomes using a statistical or machine learning model to predict outcomes as a function of covariates and treatment.

For example, to estimate the ATT we can:

1. Fit a model to predict untreated outcomes as a function of covariates and treatment
2. Use the model to predict the untreated outcome for each treated unit
3. Estimate the ATT by taking the average difference between observed and predicted outcomes for treated units

Prediction as Controlling

When we use a confounding variable to help predict counterfactual outcomes, we are “controlling” for that variable.

More formally, what these procedures are doing is *conditioning* on the confounder when computing differences in expected outcomes.

Imputation and the Common Support

Unlike matching and weighting, prediction methods do not require that the treated and untreated groups have common support in the covariates.

However, predictions outside of the common support are extrapolations (i.e. guesses) and will depend more on the modeling assumptions that the available data.

Choosing an Estimand

Imputation methods can be used to estimate a variety of average treatment effects:

- To estimate the ATT, predict counterfactuals for treated units only
- To estimate the ATU, predict counterfactuals for untreated units only
- To estimate the ATE, predict counterfactuals for all units

Regression

Statistical regression is one of the most common approaches to estimating causal effects. The basic idea is to fit a statistical model that predicts outcomes as a function of treatment and observed covariates, and then inspect the fitted model parameters to estimate treatment effects.

Regression Complications

Regression can be interpreted as a form of imputation that hides the steps of prediction and estimation inside a single procedure.

However, common regression approaches also imposes additional, non-obvious assumptions. Many people use regression without fully understanding the trade-offs they are making.

We will explore regression more deeply later in the course.

Structural Modeling

Structural modeling attempts to estimate causal effects by specifying a full model of the data generating process, including the relationships between all variables in the data.

Model parameters are estimated from the data, based on the assumption that the model is correctly specified.

Average treatment effects can then be estimated by using the model to predict counterfactual outcomes, or sometimes directly from the model parameters.

Structural modeling generally requires strong assumptions about the data generating process, but can be very powerful when the assumptions are valid.

Summary

- When treatment is not randomized, we need to worry about confounders
- We can use observed covariates to help us construct better counterfactuals and estimate causal effects
- Common approaches include matching, weighting, imputation, and structural modeling
- All approaches can be viewed as ways of conditioning on confounders to estimate better counterfactuals

A fundamental limitation

What if we can't observe all the confounding variables?