Chapter 7: Heterogeneity and Causal Moderation

A Course in Experimental Economics

John A. List

All content is derived from the Experimental Economics Textbook by John A. List (2026).

Key Ideas

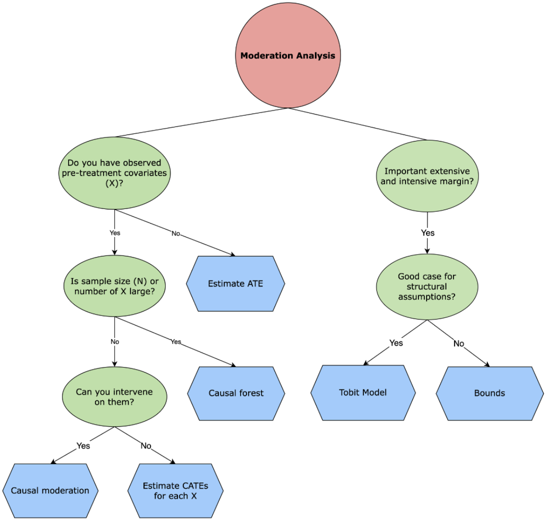
1. In the presence of potential moderation of treatment effects through predetermined covariates, researchers can estimate heterogeneous treatment effects by covariates using classical and modern tools.

2. Causal moderation analysis differs from conventional heterogeneity analysis in that the goal is to pinpoint the precise, *causal* source of the heterogeneity.

3. The limits of ex-post moderation analysis highlight the value of considering theory and identifying potential moderators in the design stage.

4. For outcomes with extensive and intensive margins, a fundamental internal validity concern can only be alleviated by invoking stronger assumptions.

Mind's Eye



Moderation Analysis
Do you have observed pre-treatment covariates (X)?
Important extensive and intensive margin?
Yes
No
Is sample size (N) or number of X large?
Estimate ATE
Yes
Good case for structural assumptions?
No
Yes
Can you intervene on them?
Causal forest
Yes
No
Tobit Model
Bounds
Yes
No
Causal moderation
Estimate CATEs for each X

Introduction

- One experimental truism in economics concerns **heterogeneity of treatment effects**
- The literature and the policymaking community are littered with facts of these types:
  - The job training program affected low income people more than high income people
  - Older workers are less responsive to health benefits than younger workers
  - The middle class is more responsive to marginal income tax rate changes than the rich
- Investigating heterogeneity tells us **how** an intervention works, by identifying the factors that influence its estimated effect
- Doing so lends insight into Experimental Problem 1 and starts to build the blocks for Experimental Problem 2

Introduction

- Fundamentally, the ideal parameter concerning heterogeneity is the variance of the individual treatment effect, $\mathbb{V}\text{ar}[\tau_i] > 0$
- Yet the common use of **between-subject** designs means we obtain information on the *marginal* distribution of potential outcomes, not the *joint* distribution
- So the analyst instead calculates average treatment effects conditional on pre-determined characteristics, or **moderators** ($\mathbf{X}_i$), yielding estimates of conditional average treatment effects

Introduction

- The **conditional average treatment effect** (CATE) for $\mathbf{X}_i = \mathbf{x}$ is defined as

$$\tau(\mathbf{x}) \equiv \mathbb{E}[\tau_i \mid \mathbf{X}_i = \mathbf{x}] = \mathbb{E}[Y_i(1) - Y_i(0) \mid \mathbf{X}_i = \mathbf{x}]$$

- The ATE of Chapter 3 can then be re-expressed as $\mathbb{E}[\tau(\mathbf{X}_i)]$, taking expectations over the distribution of $\mathbf{X}_i$

- Individuals sharing a common $\mathbf{X}_i = \mathbf{x}$ are a **subgroup**, so **subgroup average treatment effect** is used as a synonym for CATE

Introduction

- In this chapter we explore such heterogeneity, which we deem **descriptive moderation analysis**
- The analysis should not stop there: modern approaches should not only **measure** heterogeneities but also describe **why** they exist
- That is what lets us design policies and change institutions to leverage the insight
- So we also pursue **causal moderation analysis**, which invokes more restrictive assumptions but delivers a deeper understanding

Outline

Four Running Examples

Estimating Heterogeneities in Simple Cases

Basic Mechanics of Causal Moderation

Two Crucial Margins of Heterogeneity: Intensive and Extensive

Conclusions

Niederle and Vesterlund (2007)

- An innovative **lab experiment** asking whether men and women differ in their willingness to enter a competition
- Key treatment: varying the nature of the experimental pay scheme — fixed non-competitive pay versus tournament-based pay
- Along the way the authors elicit attitudes towards competition, beliefs about own and others' performance, and risk preferences as potential **moderating factors**
- A neat example of one approach to detailing *why* observed heterogeneities exist

---

Textbook section 7.1

Niederle and Vesterlund (2007)

- A substantial and significant **gender gap in tournament entry**: 73% of men and 35% of women enter
- Gender cannot be manipulated, so the authors explore the individual characteristics correlated with gender — these might suggest **mutable** causal moderators to intervene on
- Elicited beliefs account for about one-third of the gap, controlling for performance
- Their key argument turns on a competitiveness trait: "women shy away from competition and men embrace it" (p. 1067)

Dreber, Fudenberg, and Rand (2014)

- A lab experiment linking play in a **dictator game** (DG) to choices in a **repeated prisoner's dilemma** (RPD)
- Harvard undergraduates are randomly assigned to one of four versions of the RPD; the versions differ in the **returns to cooperation**
- Two main outcomes: cooperating in all rounds of the RPD, and cooperating in the first round against a given opponent
- Two secondary outcomes characterise a participant's "type":
  - *Leniency* — waiting for multiple defections before defecting
  - *Forgiveness* — reverting to cooperation once an opponent does

Dreber, Fudenberg, and Rand (2014)

- After the RPD, subjects play a DG; DG behaviour is the key dimension of heterogeneity

- The question: is DG giving a key source of **treatment effect heterogeneity** in response to cooperation incentives?

- Neither DG behaviour nor attitudes and individual predictors play an important role in predicting RPD behaviour

- Changes in the returns to cooperation do determine whether individuals cooperate — but altruism is not an important moderator

Benjamin, Choi, and Fisher (2016)

- A lab study of how priming an individual's **religious identity** affects behaviour in six classic experimental economics tasks
- We focus on the public goods game (PGG; see Chapter 2): groups of 4, each participant given a one-dollar endowment
- Outcome Y_i: the share of the endowment contributed to the group fund
- Control ($D_i = 0$): word scrambles with neutrally-framed words. Treatment ($D_i = 1$): word scrambles with religious framings
- Protestant- and agnostic/atheist-identifying individuals raise contributions by 16 and 12 cents, respectively — a role for **pre-treatment religious identity** as a causal moderator

Holz, List, A. Zentner, Cardoza, and J. E. Zentner (2023)

- A **natural field experiment** on tax compliance, in partnership with the Dominican Republic IRS
- Firms and self-employed workers randomly allocated to one of six groups
- Control: a simple reminder message before the tax deadline
- Two deterrence messages augment that baseline:
  - One raises the salience of **prison penalties** under a new law
  - One raises the salience of **public disclosure** of punishments
- Each message type is then interacted with a **commission frame** that recasts tax mistakes as voluntary choices

Holz, List, A. Zentner, Cardoza, and J. E. Zentner (2023)

- Overall, the treatments increase taxes paid by $193 million (0.23% of GDP)
- Heterogeneity is critical here because the authors observe a unique sample spanning the **entire firm size distribution**
- The top 8% of taxpayers pay over 84% of all taxes and the top 1% pay 60% — yet previous studies only reached small and medium-sized firms
- $150 million of the $193 million raised came from the top 8% of firms: heterogeneity across firm size is first order for policymakers

Essential Elements of Our Four Running Examples

Study	Type	Units	Control Condition	Treatment Condition	Outcome(s)	Key Heterogeneity Analyses
Niederle and Vesterlund (2007)	Lab	Pittsburgh Experimental Economic Lab participants	Non-competitive pay scheme	Competitive pay scheme	Choice of pay scheme and performance	Gender, beliefs, feedback aversion, preferences
Dreber, Fudenberg, and Rand (2014)	Lab	Harvard University undergraduate students	Participants in a RPD game with low returns to cooperation	Assignment to RPD game with high returns to cooperation	Various measures of cooperativeness	Whether cooperation responses to changes in cooperation incentives are stronger for those who share in the DG
Benjamin, Choi, and Fisher (2016)	Lab	Cornell University undergraduate students	Word unscrambling task with neutrally-framed words	Word unscrambling task with religiously-primed words	Many, including PGG behaviour and risk and time preference elicitations	Pre-experiment religious identity
Holz, List, A. Zentner, Cardoza, and J. E. Zentner (2023)	NFE	Individual taxpayers and firms in the Dominican Republic	Control group receiving no messaging from the tax authority	Several distinct messages, e.g. a reminder of prison-time penalties	Timing and amount of taxes paid	Response along the distribution of past and current tax payments, and by firm size

[Table 7.1 in the Textbook]

Outline

Four Running Examples

Estimating Heterogeneities in Simple Cases

- Why is the subgroup approach dominant in practice?
- Letting the Data Speak: what underpins the gender gap?
- Using Causal Forests to Estimate Heterogeneities

Basic Mechanics of Causal Moderation

Two Crucial Margins of Heterogeneity: Intensive and Extensive

Conclusions

Estimating Heterogeneities in Simple Cases

- Start with the simplest case: the dimension of heterogeneity is small and summarised by a finite collection of **subgroups**
- With large enough samples per subgroup, treat each subgroup as a **miniature randomized control trial**, holding observable covariates fixed
- Estimating $\tau(\mathbf{x})$ then reduces to a difference in means across treatment and control individuals, holding $\mathbf{x}$ fixed
- Because most analyses are guided by coarse groups (men vs. women, old vs. young) or coarse theoretical delineations (liquidity constrained or not), this is **far and away the dominant approach** in current practice

Textbook section 7.2

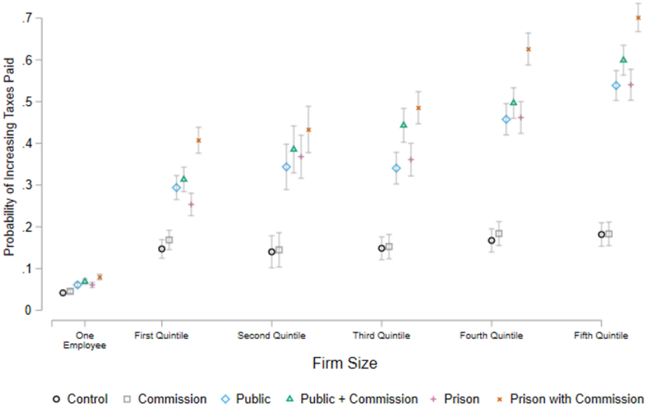
DG Giving and Cooperative Behavior in Dreber et al. (2014)

Coefficient	Pooled Estimate	DG Givers	DG Non-Givers	Difference
Treated	0.213	0.224	0.202	0.022
	(0.03)	(0.06)	(0.07)	(0.10)
Constant	0.340	0.420	0.260	
	(0.027)	(0.05)	(0.03)	
N	278	152	126	

- The difference in CATEs is small and statistically insignificant
- DG giving is **not** a key moderator for the effect of cooperation incentives on cooperative play in the RPD

[Table 7.2 in the Textbook]

Heterogeneity of Paying Taxes by Firm Size



- Holz et al. (2023): firms split into six bins — single-worker firms, then quintiles of firm size

[Figure 7.1 in the Textbook]

What the Simple Approach Delivers — and What It Does Not

- Substantial heterogeneity across firm size: the prison-with-commission treatment moves single-employee firms by 3.8 percentage points, but firms in the top size bin by **52 percentage points**
- These differences are statistically significant at conventional levels
- Dreber et al. (2014) and Holz et al. (2023) are two examples of **descriptive moderation analysis**
- Each explores a dimension of heterogeneity that speaks to the moderation aspect of Experimental Problem 1 — but this is only a first step
- Understanding *why* firm size matters is what lets us design optimal policies and change institutions

Digging into the Underpinnings of Niederle and Vesterlund (2007)

- Niederle and Vesterlund (2007) went beyond descriptive moderation to explore *why* the behavioural heterogeneity exists
- Key evidence comes from a regression model: controlling for risk preferences, overconfidence, and performance, **72% of the gender gap in tournament entry remains**
- They interpret this residual as the **competitiveness trait**
- This interpretation has been revisited by Gillen et al. (2019) and van Veldhuizen (2022)

Digging into the Underpinnings of Niederle and Vesterlund (2007)

- Neither study refutes the gender gap itself — the **descriptive moderator is consistent** across studies
- **Gillen et al. (2019):** the identification strategy is susceptible to measurement error. Measurement error in the controls understates their importance, so the residual-based approach **overstates** competitive preferences
- **van Veldhuizen (2022):** a new lab design that avoids the Gillen et al. critique explains the gap **without** a competitiveness trait — gender differences in risk attitudes, overconfidence, and performance suffice
- Building knowledge via robustness analysis and similar-population replication is a key feature of the experimental approach (Section IV of the textbook)

Using Causal Forests to Estimate Heterogeneities

- Investigations of treatment effect heterogeneity typically involve estimating a **series of CATEs**

- With many covariates this quickly becomes infeasible, and continuous covariates force functional form assumptions

- It is also hard to discover or justify ATEs conditional on interactions of multiple binary variables, or non-linear functions of continuous variables, *ex post*

- The experimental beauty suddenly loses its lustre, yielding to a series of debates about assumptions

Using Causal Forests to Estimate Heterogeneities

- Alternative: stop testing whether a particular baseline covariate is associated with different treatment effects, and instead **predict** heterogeneity using all available covariate information

- We advocate the use of **causal forests**, which adapt the classification and regression tree (CART) approach to estimating heterogeneous treatment effects

- The procedure begins by randomly splitting units into three subsamples without replacement,

$$\mathcal{J}^{train}, \quad \mathcal{J}^{estimation}, \quad \mathcal{J}^{holdout}$$

- Within $\mathcal{J}^{train}$, a regression tree-based algorithm bins observations based on their covariates

Growing a Regression Tree

- A regression tree models treatment effects as a function of covariates by **recursively partitioning** the data by each covariate
- In Holz et al. (2023), it would begin by splitting on taxes paid in the past, then on industry, and so on
- To construct a tree, the algorithm searches over all possible splits in the full training data based on a single covariate
- Each potential split forms two bins, called **leaves**
- Splits are chosen to maximise heterogeneity in treatment effects *across* leaves, subject to a penalty for the variance of treatment and control outcomes *within* a leaf

The Splitting Objective

$$\frac{1}{N^{estimation}} \sum_{\ell} \left(N_\ell \hat{\tau}_\ell^2 - 2 \left(\frac{\widehat{Var}[Y_\ell | D_i{=}1]}{\hat{\mathbb{P}}[D_i{=}1 | \mathcal{L}{=}\ell]} + \frac{\widehat{Var}[Y_\ell | D_i{=}0]}{\hat{\mathbb{P}}[D_i{=}0 | \mathcal{L}{=}\ell]} \right) \right) \tag{7.1}$$

- $N^{estimation}$: number of observations in the estimation sample
- N_ℓ: number of observations in leaf ℓ
- $\hat{\tau}_\ell$: estimated conditional average treatment effect in leaf ℓ
- $\hat{\mathbb{P}}[D_i = d \mid \mathcal{L} = \ell]$: proportion of observations in leaf ℓ from treatment $d \in \{0, 1\}$

From Splits to Estimates

- The split with the best fit is kept, generating two **child leaves** off the parent node; the algorithm then repeats within each
- It continues until too few observations remain to split, or no potential leaf improves the goodness of fit
- Given statistical independence conditional on the covariate vector, within each **terminal leaf** we estimate the difference in mean outcomes between treatment and control, for iteration b:

$$\hat{\tau}_{\ell b} = \overline{Y}_{\ell b}(1) - \overline{Y}_{\ell b}(0) \tag{7.2}$$

- $\hat{\tau}_{\ell b}$ is then assigned to all units in the full or holdout sample whose covariates place them in leaf ℓ

Averaging Across the Forest

- The researcher averages the predictions across all trees for a given unit

- That is, estimate $\hat{\tau}_i(\mathbf{X}_i)$ by averaging $\hat{\tau}_\ell$ over all B iterations of the causal forest algorithm:

$$\hat{\tau}_i(\mathbf{X}_i) = \frac{1}{B}\sum_{b=1}^{B}\hat{\tau}_{\ell b} \tag{7.3}$$

- **Dreber et al. (2014):** covariates include age and whether the participant is an economics major — cooperation under higher rewards might come from older economics majors, or perhaps young philosophy majors

- **Holz et al. (2023):** age, exemptions claimed in the previous year, sex, or marital status might moderate the treatment effect

8-Step Causal Forest Procedure (Steps 1–4)

1. Draw a subsample b from the data set without replacement containing $N_b = 0.2N$ observations.
2. Randomly split the N_b observations in half to form a training sample N^{train} and an estimation sample $N^{estimation}$. Using just the training sample:
3. For each value of each covariate $\mathbf{X}_i = \mathbf{x}$, form candidate splits of the observations into two groups based on whether $\mathbf{X}_i \leq \mathbf{x}$. Consider only splits with a minimum number of treatment and control observations in both new leaves. Choose the single split that maximizes the objective function. If the split increases the objective relative to no split, implement it and repeat this step for the new leaves. When no split can increase the objective, the current split becomes a terminal leaf.
4. Once every observation is split into a terminal leaf, the regression tree is defined for subsample b.

8-Step Causal Forest Procedure (Steps 5–8)

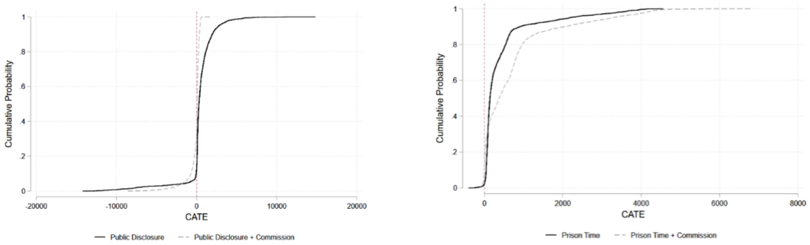
5. Using only the estimation sample, calculate $\hat{\tau}_\ell = \overline{Y}_\ell(1) - \overline{Y}_\ell(0)$ within each terminal leaf.

6. Return to the full sample of observations and assign $\hat{\tau}_{\ell b} = \hat{\tau}_\ell$ to each observation whose covariates place it in leaf ℓ, and save this prediction.

7. Repeat steps 1 through 6 a large number of times (for example, 100,000).

8. Define unit i's predicted conditional average treatment effect as $\hat{\tau}_i(\mathbf{X}_i) = \frac{1}{B} \sum_{b=1}^{B} \hat{\tau}_{\ell b}$.

CDF of CATEs in Holz et al. (2023)



[Figure 7.2 in the Textbook]

Reading the CDF of CATEs

- For a large portion of the sample, the predicted CATEs for the public disclosure and public-disclosure-with-commission messages are **zero or negligible**
- This is driven by subjects who paid no taxes in 2018 and continued to pay none in 2019
- The average effects **hide substantial heterogeneity**: about 5% of the sample reduced taxes paid relative to the previous year, some substantially
- Some units in the public disclosure treatment greatly increased taxes paid; there were no such effects once the commission frame was added
- For the prison messages, very few subjects have expected negative treatment effects, and the non-responding portion is smaller

Outline

Four Running Examples

Estimating Heterogeneities in Simple Cases

Basic Mechanics of Causal Moderation

- Why is a CATE not a causal statement about the moderator?
- Causal Moderation in Economic Experiments
- What if the moderator cannot be manipulated?

Two Crucial Margins of Heterogeneity: Intensive and Extensive

Conclusions

Setting the Stage

- Estimating CATEs is one thing; **interpreting** them is often a source of confusion

- Notation: Y_i the outcome, $D_i \in \{0, 1\}$ the treatment assignment, and $\mathbf{X}_i$ a covariate (or set of covariates) measured **prior** to treatment assignment

- With a controlled randomized experiment, randomization of D together with variation in D for each level of $\mathbf{X}_i$ affords a design that allows descriptive moderation analysis by way of CATEs:

$$\tau(\mathbf{x}) \equiv \mathbb{E}[\tau_i \mid \mathbf{X}_i = \mathbf{x}] = \mathbb{E}[Y_i(1) - Y_i(0) \mid \mathbf{X}_i = \mathbf{x}] \tag{7.4}$$

Textbook section 7.3

The Independence Condition Behind a CATE

- To recover the CATE in equation (7.4), we require a **different statistical independence condition** than in Chapter 3:

$$\{Y_i(1, X_i),\ Y_i(0, X_i)\} \perp\!\!\!\perp D_i \mid X_i$$

- In words: we require random assignment of D_i *for each* X_i

- The researcher then has a design that is internally valid for each CATE over X_i:

$$\begin{aligned}\tau(1) &= \mathbb{E}[\tau_i(1) \mid X_i = 1] = \mathbb{E}[Y_i(1,1) - Y_i(0,1) \mid X_i = 1], \\ \tau(0) &= \mathbb{E}[\tau_i(0) \mid X_i = 0] = \mathbb{E}[Y_i(1,0) - Y_i(0,0) \mid X_i = 0]\end{aligned} \tag{7.5}$$

Why We Call It *Descriptive* Moderation

- This quantity has many important applications — but it has one potential shortcoming
- We do not learn whether there is something **particular about** X_i that is generating the heterogeneity in responses ...
- ...or whether X_i is simply **correlated** with other (potentially many) covariates that are the true sources of the heterogeneity
- Returning to Holz et al. (2023): the authors obtain internally-valid CATEs for both small and large firms, but **not** an internally-valid estimate of the *causal effect of firm size* on the outcome or on the treatment effect

The Parameter We Would Like

- The counterfactual question — how does firm size *change* the treatment effect? — is a statement about how the CATE changes with X_i

$$\Delta\tau(X_i) \equiv \mathbb{E}[Y_i(1,1) - Y_i(0,1) - Y_i(1,0) - Y_i(0,0)] \qquad (7.6)$$

- This object is **not directly observable**: we observe each unit with a single potential outcome *and* with a single level of X_i

- Instead, researchers can measure the **difference between the CATEs**:

$$\Delta\tilde{\tau}(X_i) \equiv \tau(1) - \tau(0) = \mathbb{E}[\tau_i(1) \mid X_i = 1] - \mathbb{E}[\tau_i(0) \mid X_i = 0] \quad (7.7)$$

A Bias That Looks Like Selection Bias

- We can learn whether the CATE is smaller or larger when firm size changes — but not whether firm size has a **causal effect** on the treatment effect

- Add and subtract the missing counterfactual $\mathbb{E}[\tau_i(0) \mid X_i = 1]$:

$$\begin{aligned}\Delta\tilde{\tau}(X_i) = \quad & \underbrace{\mathbb{E}[\tau_i(1) - \tau_i(0) \mid X_i = 1]}_{\text{Treatment differences due to firm size}} \\ & + \underbrace{\mathbb{E}[\tau_i(0) \mid X_i = 1] - \mathbb{E}[\tau_i(0) \mid X_i = 0]}_{\text{Treatment differences due to heterogeneity}}\end{aligned} \tag{7.8}$$

- The second term is how the CATE for small firms would have differed had those units been *identical* to the large firms except for their size

Restoring Internal Validity

- To regain internal validity of $\Delta\tilde{\tau}(X_i)$ for $\Delta\tau(X_i)$, we introduce yet another distinct **statistical independence assumption**:

$$\{Y_i(1, X_i),\ Y_i(0, X_i)\} \perp\!\!\!\perp X_i \tag{7.9}$$

- This rules out differences in potential outcomes that correlate with X_i *and* predict the treatment effects

- For this to be credible, researchers should **explicitly randomize** X_i

- When that is impossible, account for all relevant differences between unit types and use **great caution** before ascribing a causal interpretation to treatment effect heterogeneity

Moderation Analysis of the PGG in Benjamin et al. (2016)

Coefficient	**Protestant-Identifying**	**Catholic-Identifying**	**Jewish-Identifying**	**Agnostic/Atheist-Identifying**
Treated	0.16	−0.18	0.01	0.12
	(0.06)	(0.07)	(0.12)	(0.06)
Constant	0.53	0.70	0.56	0.49
	(0.05)	(0.05)	(0.08)	(0.04)
N	171	138	56	168

- Groups of 4, one-dollar endowment; contributions to the group fund are doubled and split evenly

[Table 7.3 in the Textbook]

What Table 7.3 Can — and Cannot — Tell Us

- Absent any identity framing there is already heterogeneity in contribution rates: Catholics contribute more than the other three subgroups (compare the **constants** across columns)
- But selection into the experimental sample is not random, and religious identity is not explicitly randomized
- Exogenously manipulating **identity salience** is not the same as exogenously manipulating **religious affiliation**
- Two questions the design struggles with:
  - Does religious identity directly or indirectly affect PGG contributions?
  - Does religious identity directly change responsiveness to religious priming?
- Only the second falls into the realm of **causal moderation analysis**

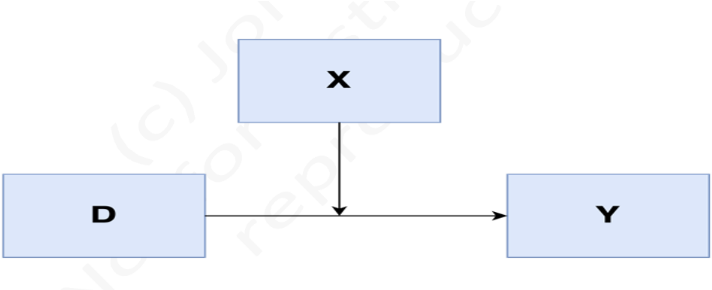
The Average Treatment Moderation Effect

- The corresponding causal moderation estimand is the **average treatment moderation effect (ATME)**:

$$ATME(x, x') \equiv \mathbb{E}\big[\{Y_i(x, 1) - Y_i(x, 0)\} - \{Y_i(x', 1) - Y_i(x', 0)\}\big] \tag{7.10}$$

- The ATME captures the **causal interaction effect**: the difference in resultant ATEs between those with $X_i = x$ and those with $X_i = x'$, when **both treatment and group are randomly assigned**

Causal Moderation Graph



[Figure 7.3 in the Textbook]

Reading the Causal Moderation Graph

- The ATME is special because it requires, at least conceptually, that X_i **can be manipulated** — moved from $X_i = x$ to $X_i = x'$ holding all else equal
- Both X and D therefore sit among the nodes that **causally precede** Y
- The node connecting D to Y is the effect we run the experiment to measure (the ATE)
- We could have drawn $X \to Y$ directly — the differential control-group contribution rates in Table 7.3 are exactly that — but we omit it to focus on moderation
- Demanding this parameter in Benjamin et al. (2016) is hard: the opportunity to randomize religion is unlikely to present itself

Designing for Causal Moderation

- Take the perspective of an experimenter planning an experiment with binary treatment D, who wants to know whether a binary covariate X^M is a **causal moderator**
- That is: does variation in X^M serve as the **direct source** of any treatment effect heterogeneity?
- Denote other covariates beyond X^M as $\mathbf{X}^{-M}$; the experimenter has a sample of N units
- We state key identifying assumptions, then ask what is needed to estimate how X^M affects outcomes separately by treatment status — **flipping the script** on the usual CATE analysis

A7.1: SUTVA with a Moderator

- Recall from Chapter 3 the full vector of treatment assignments $\mathbf{d} = (d_1, \ldots, d_N)$, giving potential outcomes $Y_i(d_i, \mathbf{d}_{-i})$

- Augmenting with X^M, the analogous notation is $Y_i(d_i, \mathbf{d}_{-i}, x_i, \mathbf{x}_{-i})$

- **Assumption 7.1 (A7.1, SUTVA-Moderator):** for all $(\mathbf{d}, \mathbf{x}) \in \{0,1\}^N \times \{0,1\}^N$,

$$Y_i(d_i, \mathbf{d}_{-i}, x_i, \mathbf{x}_{-i}) = Y_i(d_i, x_i)$$

- A7.1 requires that a unit's potential outcomes do not vary with the treatments or causal moderators assigned to, or undertaken by, **any other unit**

A7.2 and A7.3: Pre-Treatment and Conditionally Independent Moderators

- We maintain the Chapter 3 assumption that D is statistically independent of potential outcomes

- A7.2 requires that X^M and $\mathbf{X}^{-M}$ **causally precede** D; that is, the covariates are pre-treatment

- **Assumption 7.2 (A7.2, Pre-Treatment Moderators):**

$$\{Y_i(1,1), Y_i(0,1), Y_i(0,1), Y_i(0,0), X^M, \mathbf{X}^{-M}\} \perp\!\!\!\perp D_i$$

- Next, a conditional independence assumption ensuring X^M is **as-if randomly assigned**, at least conditional on $\mathbf{X}^{-M}$

- **Assumption 7.3 (A7.3, CIA Moderators):**

$$\{Y_i(1,1), Y_i(0,1), Y_i(0,1), Y_i(0,0)\} \perp\!\!\!\perp X^M \mid \mathbf{X}^{-M}$$

A7.4: Common Support

- A7.3 may look strong or undesirable — but where we **can** explicitly randomize X^M, it is exactly akin to our assumption on the assignment mechanism for internally-valid recovery of the ATE

- Finally, we need the practical condition that there is **variation in** X^M at the level at which we estimate its causal contribution

- **Assumption 7.4 (A7.4, Moderator Common Support):**

$$\mathbb{P}[X_i^M = 1 \mid \mathbf{X}_i^{-M}] \in (0, 1)$$

Recovering the ATME

With A7.1–A7.4 in place, together with randomized assignment of treatment:

$$
\begin{aligned}
ATME(1,0) &\equiv \mathbb{E}\big[\{Y_i(1,1) - Y_i(1,0)\} - \{Y_i(0,1) - Y_i(0,0)\}\big] \\
&= \mathbb{E}_{\mathbf{X}^{-M}}\big[\mathbb{E}[Y_i \mid D_i = 1, X_i^M = 1, \mathbf{X}^{-M}] \\
&\quad - \mathbb{E}[Y_i \mid D_i = 0, X_i^M = 1, \mathbf{X}^{-M}]\big] \\
&\quad - \mathbb{E}_{\mathbf{X}^{-M}}\big[\mathbb{E}[Y_i \mid D_i = 1, X_i^M = 0, \mathbf{X}^{-M}] \\
&\quad - \mathbb{E}[Y_i \mid D_i = 0, X_i^M = 0, \mathbf{X}^{-M}]\big]
\end{aligned}
\tag{7.11}
$$

A Difference-in-Differences of CATEs

- The intuition: use the CIA for moderators to recover the mean of each of the **four** potential outcomes in the ATME, then take the corresponding difference in differences

- Where X^M **is** randomly assigned — our motivating research design — we get an unconditional version of A7.3 for free, and (7.11) simplifies:

$$
\begin{aligned}
ATME(1,0) = &\left[\mathbb{E}[Y_i \mid D_i = 1, X_i^M = 1] - \mathbb{E}[Y_i \mid D_i = 0, X_i^M = 1]\right] \\
&- \left[\mathbb{E}[Y_i \mid D_i = 1, X_i^M = 0] - \mathbb{E}[Y_i \mid D_i = 0, X_i^M = 0]\right]
\end{aligned}
\tag{7.12}
$$

- If religious identity were as good as randomly assigned in Benjamin et al. (2016), Table 7.3 would say that receiving and maintaining Catholicism rather than Protestantism **lowers PGG contributions by 34 cents**

Immutable Moderators

- So far the moderator has been, at least conceptually, **manipulable**: something the experimenter could vary holding all else constant
- But what about moderators without this property?
- Can we still make progress on the causal sources of heterogeneity in **age, race, gender, and even sex**?
- This is precisely the nature of the query in Niederle and Vesterlund (2007)

The Niederle and Vesterlund (2007) Design

- Subjects completed a standard experimental economics task — adding up two-digit numbers
- In the initial rounds they participated in **both** noncompetitive and competitive pay schemes:
  - Noncompetitive: piece rate of 50 cents per correctly-added number
  - Tournament: matched with other participants, $2 per correct problem if they solve the most in their tournament, zero otherwise
- After these rounds, and **without additional feedback**, subjects chose which pay scheme to face in subsequent rounds

Behind Immutable Characteristics Lie Mutable Features

- Descriptive moderation: 73% of men and 35% of women enter the tournament
- Going beyond description, the authors explore individual characteristics correlated with gender
- Their data point to **preferences, beliefs, and constraints** — and some of these factors are **mutable**
- The lesson in stark relief: behind immutable characteristics may lie opportunities to study the causal moderators explicitly
- Providing the underpinnings of CATEs is invaluable for a policymaker or decisionmaker interested in driving change

Why are Women More Risk Averse in Their Mutual Fund Investments?

- A stylized result: women are more risk averse than men (Croson and Gneezy, 2009)
- So when Dwyer et al. (2002) reported that women take less risk in their most recent, largest, and riskiest mutual fund investments, few were surprised
- But using a national survey of nearly 2,000 mutual fund investors, they also report that the impact of gender **weakens significantly** once investor knowledge of financial markets is controlled for
- The gender heterogeneity can be **substantially explained by knowledge disparities** — and knowledge can be targeted
- The practical implication: enhance educational investment marketing efforts that target women

Outline

Four Running Examples

Estimating Heterogeneities in Simple Cases

Basic Mechanics of Causal Moderation

Two Crucial Margins of Heterogeneity: Intensive and Extensive
- How do scientists decompose an ATE into margins?
- Bounding the Intensive and Extensive Margin Effects
- Using Baseline Outcome Data to Identify Intensive Margin Effects
- A Tobit Approach to Estimating Margins

Conclusions

Heterogeneity Is Lurking on Every Margin

- In Dreber et al. (2014), 14–29% of subjects never cooperated in the RPD and 55% give nothing in the DG
- In Holz et al. (2023), **70% of control group firms paid no taxes in 2019**
- Researchers therefore decompose the ATE into:
  - the part attributable to inducing a unit to move from a zero to a positive outcome — **the extensive margin**
  - the part attributable to increasing the outcome for those who already have a positive outcome — **the intensive margin**
- These questions become first order for job training and labour supply, welfare programs and health outcomes, or tax law changes and charitable giving

Textbook section 7.4

The ATE Is Identified; The Margins Are Not

- Because the amount of taxes paid is a **non-negative** random variable, the expected amount paid can be written as

$$\mathbb{E}[Y_i \mid D_i] = \mathbb{P}[Y_i > 0 \mid D_i]\, \mathbb{E}[Y_i \mid Y_i > 0, D_i] \tag{7.13}$$

- Using the assumptions from Chapter 3, the difference in treatment and control means decomposes as:

$$\begin{aligned}
\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] &= \mathbb{E}[Y_i(1) - Y_i(0)] \\
&= \underbrace{\big(\mathbb{P}[Y_i(1) > 0] - \mathbb{P}[Y_i(0) > 0]\big)}_{\text{Extensive Margin Effect}}\, \mathbb{E}[Y_i(1) \mid Y_i(1) > 0] \\
&\quad + \underbrace{\big(\mathbb{E}[Y_i(1) \mid Y_i(1) > 0] - \mathbb{E}[Y_i(0) \mid Y_i(0) > 0]\big)}_{\text{Intensive Margin Effect}}\, \mathbb{P}[Y_i(0) > 0]
\end{aligned} \tag{7.14}$$

Where the Selection Bias Comes From

- The first term is the **extensive margin effect**: the fraction of taxpayers in treatment who paid taxes versus the fraction in control. It **is** identified under the Chapter 3 assumptions

- The **intensive margin effect** — the difference in tax payments conditional on paying positive taxes — is **not**:

$$\begin{aligned}\mathbb{E}[Y_i(1) \mid Y_i(1) > 0] - \mathbb{E}[Y_i(0) \mid Y_i(0) > 0] \\ = \underbrace{\mathbb{E}[Y_i(1) - Y_i(0) \mid Y_i(1) > 0]}_{\text{Causal Intensive Margin Effect}} \\ + \underbrace{\mathbb{E}[Y_i(0) \mid Y_i(1) > 0] - \mathbb{E}[Y_i(0) \mid Y_i(0) > 0]}_{\text{Selection Bias}}\end{aligned} \tag{7.15}$$

- The bias emerges because treatment **changes the composition** of taxpayers who pay positive taxes

Stratifying Units into Types

- As in Chapters 3 and 12 (types of attritors) and Chapters 3 and 13 (types of compliers), we begin by **stratifying units into types**

- Four exhaustive types, based on potential outcomes:
  - **Always-Filers** — file taxes regardless of assignment
  - **Never Filers** — do not file regardless of assignment
  - **Control Switchers** — would not file if treated, but would file if assigned to control
  - **Treatment Switchers** — would not file if assigned to control, but would if treated

Types of Units with Extensive and Intensive Margins of Response

	$Y_i(1) = 0$	$Y_i(1) > 0$
$Y_i(0) = 0$	Never Filers	Treatment Switchers
$Y_i(0) > 0$	Control Switchers	Always Filers

[Table 7.4 in the Textbook]

Understanding the Internal Validity Issues

- Using this stratification, we can understand the **internal validity issues** that arise in the decomposition
- The **intensive margin effect** is the effect of the treatment on the **always filers**
- In contrast, the **extensive margin effect** is the effect of the treatment on the **control switchers** and the **treatment switchers**
- This means that we can write the decomposition of the ATE based on the **joint distribution of potential outcomes** as

$$
\begin{aligned}
&\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] = \mathbb{E}[Y_i(1) - Y_i(0) \mid Y_i(1), Y_i(0)] \\
&= \mathbb{E}[Y_i(1) \mid Y_i(0) = 0, Y_i(1) > 0]\, \mathbb{P}[Y_i(0) = 0, Y_i(1) > 0] \\
&- \mathbb{E}[Y_i(0) \mid Y_i(0) > 0, Y_i(1) = 0]\, \mathbb{P}[Y_i(0) > 0, Y_i(1) = 0] \\
&+ \mathbb{E}[Y_i(1) - Y_i(0) \mid Y_i(0) > 0, Y_i(1) > 0]\, \mathbb{P}[Y_i(0) > 0, Y_i(1) > 0]
\end{aligned} \tag{7.16}
$$

What the Decomposition Reveals

- The effect on the **never filers** is always zero
- The first two terms are the **extensive margin effects** on the overall tax receipts — that is, the effects for the control-only switchers and the treatment-only switchers. The third term is the **intensive margin effect**
- Under this decomposition, it is easy to see that our experimental approach **does not yield causal quantities**: when decomposing the effects, we are comparing **different subpopulations** — control switchers, treatment switchers, and always filers
- The fundamental problem is that these are **different populations**, so it is not possible to attribute fractions of the tax revenue ATE to $\tau^{Extensive}$ and $\tau^{Intensive}$
- With this understanding in hand, we can begin to impose restrictions that allow us to **bound** the treatment effects

Bounding the Two Margins

- A key assumption is **monotonicity**, also called **monotone treatment response**: it rules out that both control switchers and treatment switchers can simultaneously exist, and to elucidate the approach we rule out the existence of **control-only switchers**
- That is, no person would have paid positive taxes in the control group but would not have paid positive taxes in the treatment group: $Y_i(1) - Y_i(0) \geq 0 \; \forall i$, which implies a **positive ATE**
- Under this assumption we can bound the treatment effects as

$$\tau^{Extensive} \in \left(0, \; \frac{\tau}{\mathbb{P}[Y_i(0) = 0, Y_i(1) > 0]}\right) \tag{7.19a}$$

$$\tau^{Intensive} \in \left(0, \; \frac{\tau}{\mathbb{P}[Y_i(0) > 0, Y_i(1) > 0]}\right) \tag{7.19b}$$

- Equations 7.19a and 7.19b provide bounds which the analyst can use to determine the relevant contributions of both the extensive and the intensive margins driving enhanced tax receipts due to treatment

Derivation

A7.5 and A7.6

- An alternative to bounding: collect **baseline outcome data**, which recovers the intensive margin under three additional assumptions

- **Assumption 7.5 (A7.5, No Pre-treatment Effect):**

$$\mathbb{E}[Y_{i0}(1) - Y_{i0}(0) \mid Y_{i1} > 0, Y_{i0} > 0] = 0$$

  - The treatment effect in the pre-treatment period is zero; under randomization this holds by definition
  - But one may want to stratify on an indicator for whether the outcome is positive in the pre-period

- **Assumption 7.6 (A7.6, Common Trends):**

$$\begin{aligned}&\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid Y_{i1}(1) > 0, Y_{i0}(1) > 0, D_i = 1]\\ &\quad = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid Y_{i1}(0) > 0, Y_{i0}(0) > 0, D_i = 0]\end{aligned}$$

A7.7 and the Intensive Margin Effect

- A7.6 differs from the standard common trends assumption of the difference-in-differences design: the trend must hold in the *subsample* with positive outcomes in **both** the baseline and follow-up periods
- **Assumption 7.7 (A7.7, Time Monotonicity at the Extensive Margin):**

$$Y_{i1}(d) > 0 \Rightarrow Y_{i0}(d) > 0 \quad \forall\, i, d$$

  - No units with a positive outcome at follow-up who lack one at baseline
  - Rules out time trends affecting the extensive margin decision
  - Strong; usually holds only when a positive outcome is an **absorbing state** — so it is unlikely to hold in Holz et al. (2023)
- With A7.5–A7.7 the intensive margin effect is identified by **difference-in-differences**:

$$\begin{aligned}\tau^{Intensive} &= \mathbb{E}[Y_{i1}(1) - Y_{i0}(1) \mid Y_{i1} > 0, Y_{i0} > 0, D_i = 1] \\ &\quad - \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid Y_{i1} > 0, Y_{i0} > 0, D_i = 0]\end{aligned} \tag{7.20}$$

A Tobit Approach

- As an alternative to the approaches above, the analyst can estimate a **Tobit model**

- A Tobit model assumes the following data generating process:

$$Y_i = \max\{0,\ \beta_0 + \beta_1 D_i + U_i\}, \qquad U_i \sim N(0, \sigma^2) \tag{7.21}$$

- Using this approach, we can estimate the desired parameters, as summarized in Table 7.5

Decomposition of Effects in a Tobit Model

	Tobit Model
Population fractions of groups in decomposition	
$\mathbb{P}[Y_i(0) = 0, Y_i(1) > 0]$	$\Phi\left(\frac{\beta_0+\beta_1}{\sigma}\right) - \Phi\left(\frac{\beta_0}{\sigma}\right)$
$\mathbb{P}[Y_i(0) > 0, Y_i(1) > 0]$	$\Phi\left(\frac{\beta_0}{\sigma}\right)$
Group-specific ATE in causal decomposition	
$\tau^{Extensive}$	$\beta_0 + \beta_1 + \sigma \dfrac{\phi\left(\frac{\beta_0+\beta_1}{\sigma}\right) - \phi\left(\frac{\beta_0}{\sigma}\right)}{\Phi\left(\frac{\beta_0+\beta_1}{\sigma}\right) - \Phi\left(\frac{\beta_0}{\sigma}\right)}$
$\tau^{Intensive}$	β_1

[Table 7.5 in the Textbook]

The Cost of the Tobit Shortcut

- The Tobit approach is straightforward, but it requires **strong assumptions** about the data generating process
- The main issue: the connection between the latent-model parameters and the causal effects requires certain **distributional** assumptions
- When those are not satisfied, estimates can depart greatly from the truth
- For example, **heteroskedasticity or non-normality** result in inconsistent estimates

Giving at the Opera: Both Intensively and Extensively

- Huck and Rasul (2011) partnered with the Bavarian State Opera House to test **matched donations**, sending 14,000 mailers for a youth program
- Four conditions: control; told the project had raised 60,000 euros but needed more; donations matched 0.5:1 up to 60,000 euros; donations matched 1:1 up to 60,000 euros
- Theory sharpens the margins:
  - **Pure altruism** $\Rightarrow$ donors target an amount, so matching **crowds out** donations
  - **Pure warm glow** $\Rightarrow$ matching **crowds in** donations
  - **Mixed motives** $\Rightarrow$ partial crowding out
- To test the relevant theories, Huck and Rasul (2011) had to be able to identify different changes along the **extensive** and **intensive** margins

Outline

Four Running Examples

Estimating Heterogeneities in Simple Cases

Basic Mechanics of Causal Moderation

Two Crucial Margins of Heterogeneity: Intensive and Extensive

Conclusions

What have we learned?

- **Heterogeneity is ubiquitous.** Combined with the growth in subject-level information, this has driven rapid growth in applying the experimental method to measure it
- Since $\mathbb{V}\text{ar}[\tau_i] > 0$ drives researchers to identify differences in treatment effects across subgroups, **recovery of CATEs has become a central goal**
- Describing individual differences in preferences, beliefs, and constraints remains a key scientific pursuit
- We summarised the tools available to estimate these objects **as causal parameters** — and the pitfalls researchers may face

Textbook section 7.5

What have we learned?

- For outcomes with both margins, the contribution from extensive and intensive margin effects is not identified without monotonicity, baseline data, or distributional assumptions

- While exploring moderation helps us to understand what exogenous variables affect the strength and direction of the causal relationship of interest, such lessons represent only the **scientific beginning** of how our experimental toolkit can help us understand the world around us

- **Next chapter:** we consider a distinct framework, **mediation analysis**, which seeks to answer a different question — what is the causal chain, or mechanism, through which treatments yield effects?

- The combination of moderation and mediation not only generates knowledge into **EP1** but also begins to build toward deepening insights into **EP2**

References I

Athey, S. and G. Imbens (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences 113*(27), 7353–7360.

Benjamin, D. J., J. J. Choi, and G. Fisher (2016). Religious identity and economic behavior. *The Review of Economics and Statistics 98*(4), 617–637.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernández-Val (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India. NBER Working Paper 24678, National Bureau of Economic Research.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic Literature 47*(2), 448–474.

Dreber, A., D. Fudenberg, and D. G. Rand (2014). Who cooperates in repeated games: The role of altruism, inequity aversion, and demographics. *Journal of Economic Behavior & Organization 98*, 41–55.

Dwyer, P. D., J. H. Gilkeson, and J. A. List (2002). Gender differences in revealed risk taking: Evidence from mutual fund investors. *Economics Letters 76*(2), 151–158.

Gillen, B., E. Snowberg, and L. Yariv (2019). Experimenting with measurement error: Techniques with applications to the Caltech cohort study. *Journal of Political Economy 127*(4), 1826–1863.

Holz, J. E., J. A. List, A. Zentner, M. Cardoza, and J. E. Zentner (2023). The $100 million nudge: Increasing tax compliance of firms using a natural field experiment. *Journal of Public Economics 218*, 104779.

Huck, S. and I. Rasul (2011). Matched fundraising: Evidence from a natural field experiment. *Journal of Public Economics 95*(5), 351–362. Charitable Giving and Fundraising Special Issue.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Staub, K. E. (2014). A causal interpretation of extensive and intensive margin effects in generalized Tobit models. *The Review of Economics and Statistics 96*(2), 371–375.

van Veldhuizen, R. (2022). Gender differences in tournament choices: Risk preferences, overconfidence, or competitiveness? *Journal of the European Economic Association 20*(4), 1595–1618.

Wager, S. and S. Athey (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association 113*(523), 1228–1242.

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*, Volume 1. Cambridge, MA: The MIT Press.

Bounding the Margins: The Decomposition (1/4)

The intensive margin effect is the effect of the treatment on the always filers. In contrast, the extensive margin effect is the effect of the treatment on the control switchers and the treatment switchers. This means that we can write the decomposition of the ATE based on the joint distribution of potential outcomes as

$$\begin{aligned}
&\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] = \mathbb{E}[Y_i(1) - Y_i(0) \mid Y_i(1), Y_i(0)] \\
&= \mathbb{E}[Y_i(1) \mid Y_i(0) = 0, Y_i(1) > 0]\,\mathbb{P}[Y_i(0) = 0, Y_i(1) > 0] \\
&- \mathbb{E}[Y_i(0) \mid Y_i(0) > 0, Y_i(1) = 0]\,\mathbb{P}[Y_i(0) > 0, Y_i(1) = 0] \\
&+ \mathbb{E}[Y_i(1) - Y_i(0) \mid Y_i(0) > 0, Y_i(1) > 0]\,\mathbb{P}[Y_i(0) > 0, Y_i(1) > 0]
\end{aligned} \tag{7.16}$$

- The effect on the never filers is always zero
- The first two terms are the extensive margin effects on the overall tax receipts; the third term is the intensive margin effect

Back to Main Slides

Bounding the Margins: Imposing Monotonicity (2/4)

A key assumption is monotonicity, also called monotone treatment response. This assumption rules out that both control switchers and treatment switchers can simultaneously exist. To elucidate the approach, we rule out the existence of control-only switchers: we assume that no person would have paid positive taxes in the control group but would not have paid positive taxes in the treatment group. This assumption can be written as $Y_i(1) - Y_i(0) \geq 0 \; \forall i$; notice that this assumption implies a positive ATE. Under this assumption, we can write the ATE as

$$\begin{aligned}
\mathbb{E}[Y_i(1) - Y_i(0)] &= \mathbb{E}[Y_i(1) \mid Y_i(0) = 0, Y_i(1) > 0] \, \mathbb{P}[Y_i(0) = 0, Y_i(1) > 0] \\
&\quad + \big(\mathbb{E}[Y_i(1) \mid Y_i(0) > 0, Y_i(1) > 0] \\
&\qquad - \mathbb{E}[Y_i(0) \mid Y_i(0) > 0, Y_i(1) > 0]\big) \, \mathbb{P}[Y_i(0) > 0, Y_i(1) > 0] \\
&= \mathbb{P}[Y_i(0) = 0, Y_i(1) > 0] \, \tau^{Extensive} + \mathbb{P}[Y_i(0) > 0, Y_i(1) > 0] \, \tau^{Intensive}
\end{aligned} \tag{7.17}$$

Bounding the Margins: What the Data Identify (3/4)

Under monotonicity, we can identify three of these terms from the data:

$$
\begin{aligned}
\mathbb{P}[Y_i(0) = 0, Y_i(1) > 0] &= \mathbb{P}[Y_i > 0 \mid D_i = 1] - \mathbb{P}[Y_i > 0 \mid D_i = 0] \\
\mathbb{P}[Y_i(0) > 0, Y_i(1) > 0] &= \mathbb{P}[Y_i > 0 \mid D_i = 0] \\
\mathbb{E}[Y_i(0) \mid Y_i(0) > 0, Y_i(1) > 0] &= \mathbb{E}[Y_i \mid Y_i > 0, D_i = 0]
\end{aligned}
\tag{7.18}
$$

- However, we cannot identify $\mathbb{E}[Y_i(1) \mid Y_i(0) = 0, Y_i(1) > 0]$ or $\mathbb{E}[Y_i(1) \mid Y_i(0) > 0, Y_i(1) > 0]$
- This means that it is not possible to attribute fractions of the tax revenue ATE to $\tau^{Extensive}$ and $\tau^{Intensive}$

Bounding the Margins: The Bounds (4/4)

However, we can bound these effects. To see how, first notice that the monotonicity assumption implies a positive ATE, which means that the domains of $\tau^{Extensive}$ and $\tau^{Intensive}$ are $\mathbb{R}^+$. Using the identified probabilities from equation 7.18 and the decomposition in equation 7.17, along with the constraint that both effects are positive ($\mathbb{R}^+$), we can bound the treatment effects as

$$\tau^{Extensive} \in \left(0, \ \frac{\tau}{\mathbb{P}[Y_i(0) = 0, Y_i(1) > 0]}\right) \tag{7.19a}$$

$$\tau^{Intensive} \in \left(0, \ \frac{\tau}{\mathbb{P}[Y_i(0) > 0, Y_i(1) > 0]}\right) \tag{7.19b}$$

Equations 7.19a and 7.19b provide bounds for which the analyst can use to determine the relevant contributions of both the extensive and the intensive margins driving enhanced tax receipts due to treatment.