Causality

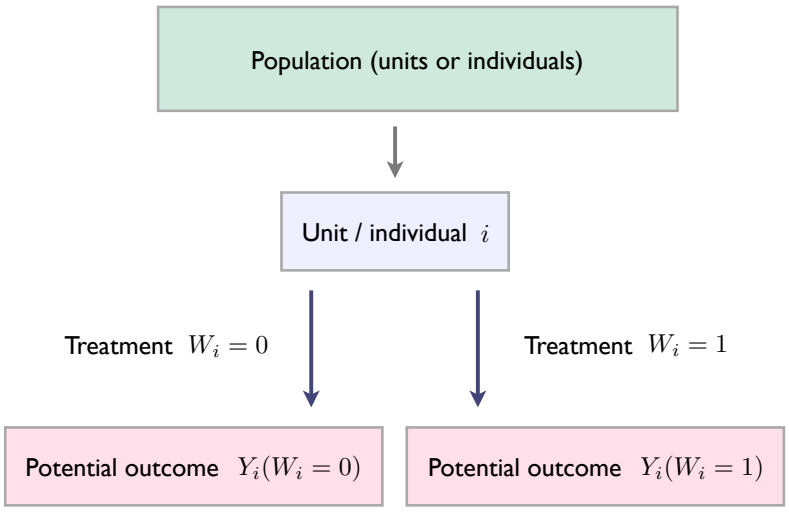
Data Science for Marketing Decision Making
Günter J. Hitsch
Chicago Booth

Winter 2017

Overview

1. What is causality? — The potential outcomes model
2. Estimating average treatment effects
   - Randomized controlled trials (A/B testing)
   - Observational studies and natural experiments
3. Causality and regression analysis
4. Application: Measuring income and wealth effects on private label demand

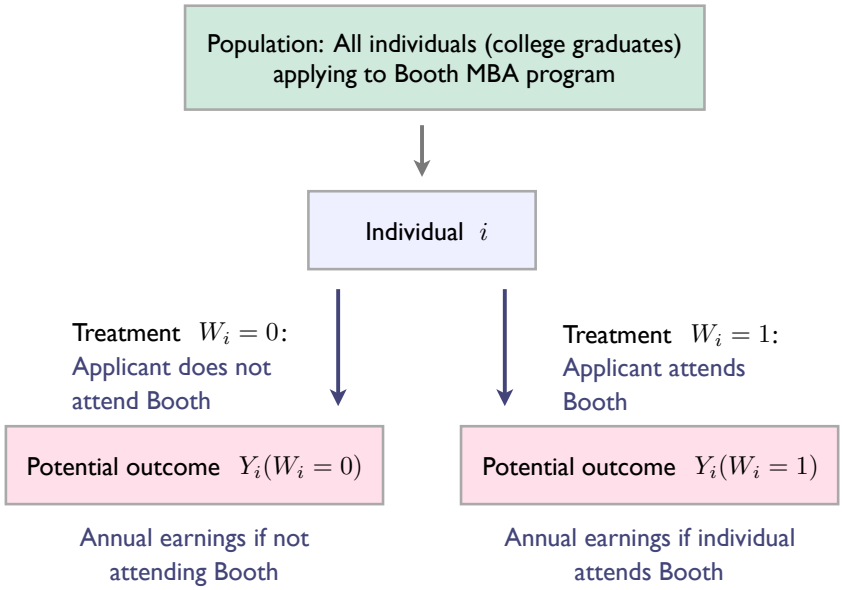
The potential outcomes model



Populations are groups of units
(consumers, patients, stores, markets, …)

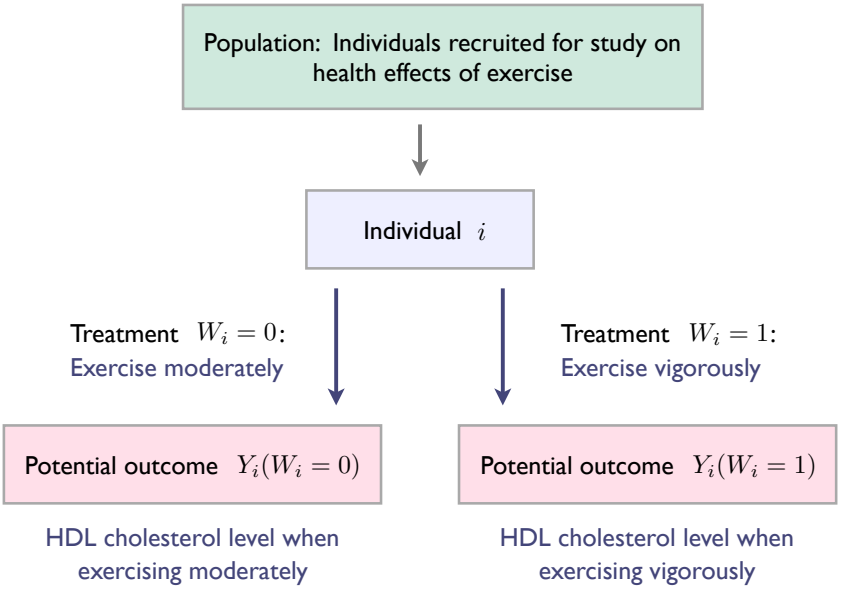
Two treatment levels: 0, 1
Can be generalized to multiple treatments

Example: Attending Chicago Booth MBA program and earnings



At age 50

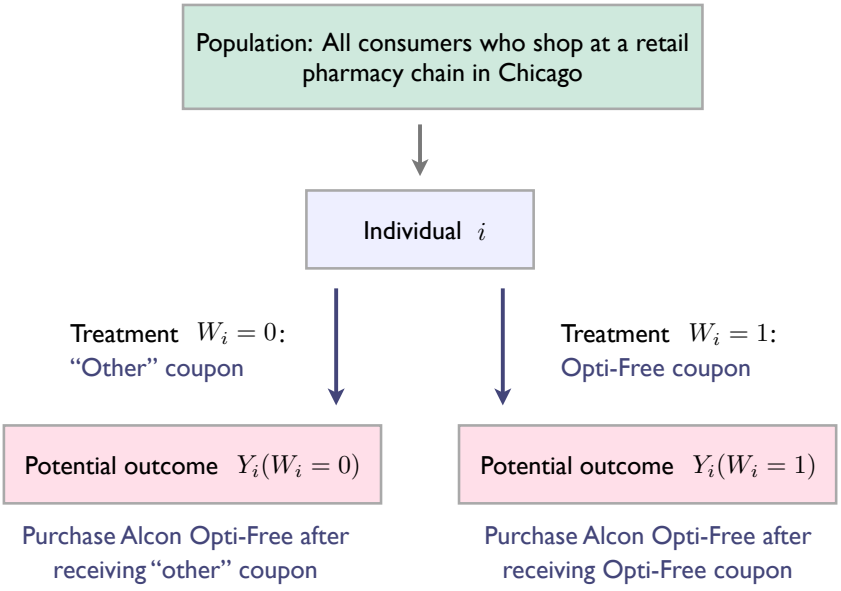
Example: Exercise intensity and HDL cholesterol levels



6 months period

HDL: high-density lipoprotein

Example: Receiving a target coupon and purchase rate



Receive target coupon for Alcon Opti-Free contact lens solution or coupon for product in different product category at check-out

Purchase within 10 weeks of receiving the coupon

Causal effect (treatment effect)

- **Causal effect (treatment effect)** of the treatment W_i on individual i:

$$\underbrace{Y_i(W_i = 1) - Y_i(W_i = 0)}_{\text{Difference in potential outcomes}}$$

- Potential outcomes framework is a model of parallel worlds, where the treatment is the only (!) difference across the two worlds

- Difference in earnings when attending vs. not attending Chicago Booth
- Difference in HDL cholesterol level when exercising vigorously vs. exercising moderately
- Difference in purchase rate when receiving vs. not receiving Opti-Free target coupon

Potential outcomes vs. actually observed data

- In the data we only observe the *realized outcome* for each individual:

$$Y_i^{obs} = \begin{cases} Y_i(W_i = 0) & \text{if } W_i = 0 \text{ HDL level if individual exercises moderately} \\ Y_i(W_i = 1) & \text{if } W_i = 1 \text{ HDL level if individual exercises vigorously} \end{cases}$$

  - Only observe one, never both outcomes
- This means we cannot directly estimate the individual causal effect (treatment effect)

$$Y_i(W_i = 1) - Y_i(W_i = 0)$$

  - One of the potential outcomes to estimate the causal effect is not in the data

Average treatment effect

- Instead of trying to estimate individual treatment effects, maybe we can estimate the **average treatment effect** in the population:

$$\text{ATE} = \mathbb{E}[Y_i(W_i = 1) - Y_i(W_i = 0)]$$

- *Average difference* in earnings when attending vs. not attending Chicago Booth
- *Average difference* in HDL cholesterol level when exercising vigorously vs. exercising moderately
- *Average difference* in purchase rate when receiving vs. not receiving Opti-Free target coupon

Estimating the average treatment effect

- Remember what data we observe:

$$Y_i^{obs} = \begin{cases} Y_i(W_i = 0) & \text{if } W_i = 0 \\ Y_i(W_i = 1) & \text{if } W_i = 1 \end{cases}$$

- Can we estimate the ATE, $\mathbb{E}[Y_i(W_i = 1) - Y_i(W_i = 0)]$, using the observed data?

$$\tau = \mathbb{E}[Y_i^{obs} | W_i = 1] - \mathbb{E}[Y_i^{obs} | W_i = 0]$$

Attempt to estimate the average treatment effect

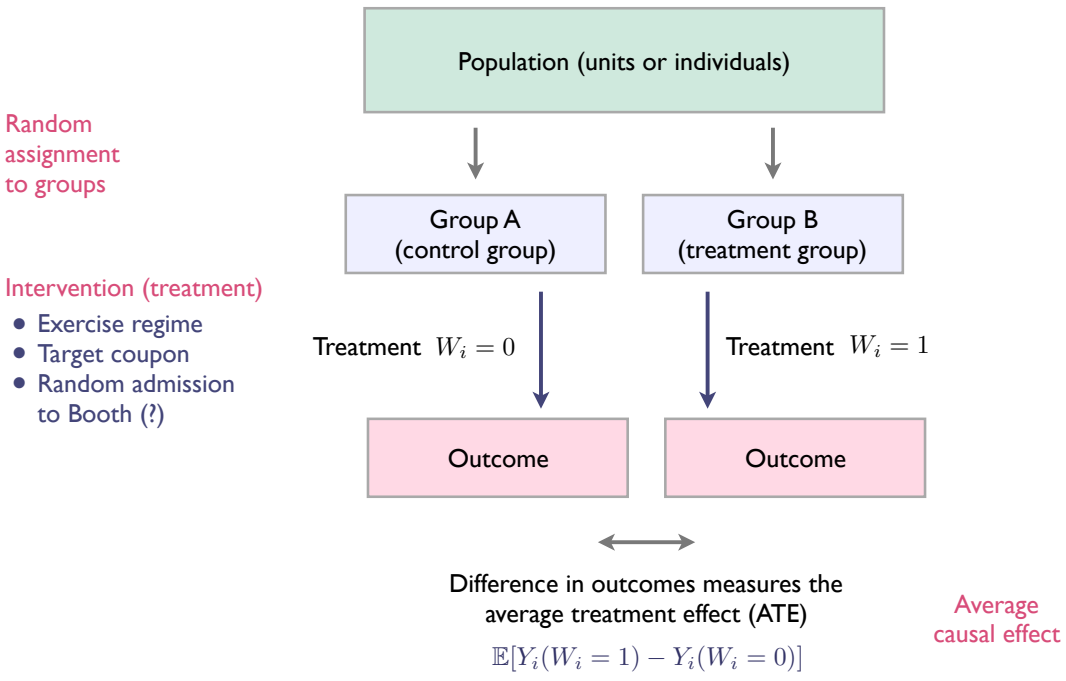
- Average difference in earnings among applicants who attended or did not attend Chicago Booth
- Average difference in HDL cholesterol levels among individuals who exercised vigorously or exercised moderately
- Average difference in purchase rates among consumers who received or did not receive an Opti-Free target coupon

Confounds

- Confounds — "*correlation does not imply causation*"
  - Individuals who are admitted to Booth may have better skills and a higher earnings potential than those who do not attend Booth
  - Individuals who exercise vigorously may generally be healthier (allowing them to exercise vigorously) than those who exercise moderately
  - Suppose Alcon Opti-Free coupon is targeted to individuals who buy the product at the current purchase occasion. Then individuals with a coupon for Alcon Opti-Free will generally be more likely to buy in future than those who receive a different coupon.

- In general, τ is not the average treatment effect (ATE)
- We never observe both parallel worlds, and hence *cannot generally estimate the average treatment effect from observed data*

Randomized controlled trials (RCT's) — A/B testing



Multiple treatments

- Note: Approach can be generalized
- Create groups A, B, C, D, ... with multiple treatments

Randomization to estimate the average treatment effect

Key conditions:

1. **Unconfoundedness**: The potential outcomes, $Y_i(0)$ and $Y_i(1)$ are statistically independent of the treatment W_i
2. **Overlap**: The probability of receiving the treatment is neither 0 nor 1:
$$0 < \Pr\{W_i = 1\} < 1$$
3. **SUTVA** (**stable unit treatment value assumption**): The treatment assignment for one unit i does not affect the outcome for a different unit, $k \neq i$

1. and 2. are under the experimenter's control: The treatment is chosen randomly (say by flipping a coin)

3. Rules out social interactions or economic equilibrium effects.

If conditions 1.-3. are satisfied, then

$$\tau = \mathbb{E}[Y_i^{\text{obs}}|W_i = 1] - \mathbb{E}[Y_i^{\text{obs}}|W_i = 0] = \text{ATE}$$

Making the argument mathematically tight

Below, Y_i is short-hand for Y_i^{obs} — the observed outcome.

$$\begin{aligned}\mathbb{E}[Y_i(1) - Y_i(0)] &= \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)] \\ &= \mathbb{E}[Y_i(1)|W_i = 1] - \mathbb{E}[Y_i(0)|W_i = 0] \\ &= \mathbb{E}[Y_i|W_i = 1] - \mathbb{E}[Y_i|W_i = 0]\end{aligned}$$

The first line follows from the linearity of the expectation operator.

The second line follows from the *unconfoundedness* assumption: W_i and $Y_i(0), Y_i(1)$ are independent. This is crucial to see why randomization allows us to infer the ATE. In the MBA admission example we suspect that $\mathbb{E}[Y_i(1)|W_i = 1] > \mathbb{E}[Y_i(1)]$. This is called *selection* (selection bias, selection effect).

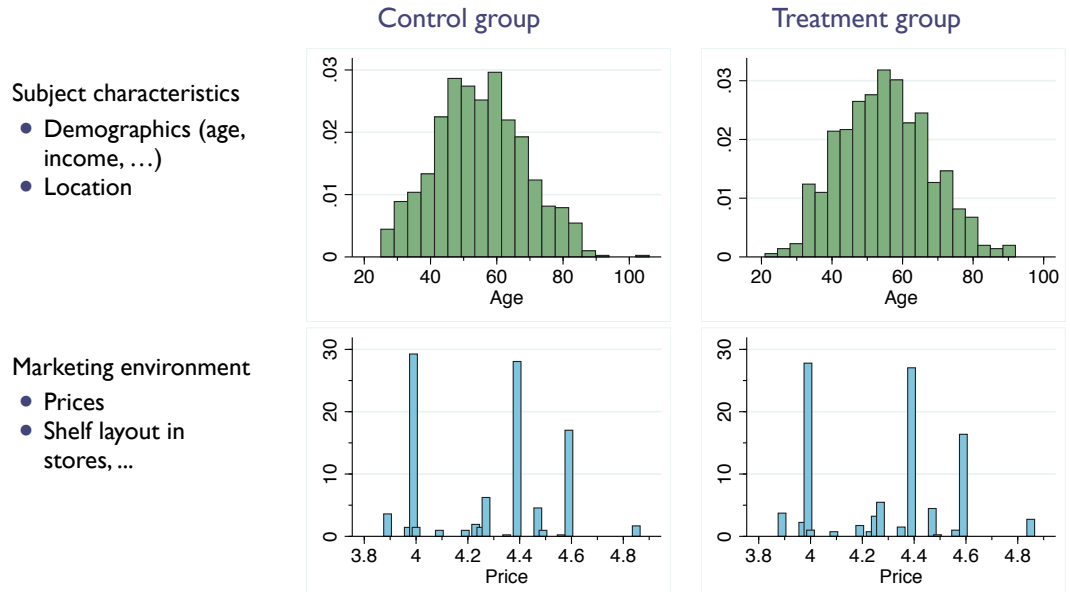
The third line says that if individual i is assigned treatment $w = 0, 1$ then the observed outcome Y_i is the same as the potential outcome $Y_i(w)$ corresponding to this treatment.

Intuition

- Random assignment creates *two replicas of the same population of units*
- *Only difference* between the two populations is the treatment assignment — the experimental *manipulation*
- Implies that the difference in outcomes must be *caused* by the treatment

Successful design of an experiment (RCT's or A/B testing)

- Randomization → control and treatment groups are replicas of the same population



Test if control and treatment group represent the same population

- Eyeballing
- Formal statistical tests: t-test for equality of means, Kolmogorov-Smirnov test, ...

```
rct_DT[, mean(age), by = treatment_indicator]

   treatment_indicator       V1
1:                   1 55.86165
2:                   0 55.25467

t.test(age ~ treatment_indicator, data = rct_DT)


Welch Two Sample t-test

data:  age by treatment_indicator
t = -1.0232, df = 1997.8, p-value = 0.3063
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.7703775  0.5564226
sample estimates:
mean in group 0 mean in group 1
       55.25467        55.86165
```

Estimating the average treatment effect using regression

Note: The treatment indicator $W_i = 0, 1$ is a dummy variable

Regression model to estimate the average treatment effect τ:

$$Y_i = \beta_0 + \tau W_i + (\beta_1 X_{i1} + \cdots + \beta_k X_{iK}) + \epsilon_i$$

- $X_{i1}, \ldots, X_{iK}$ are called *pre-treatment* variables

Do we need to control for the pre-treatment variables $X_{i1}, \ldots, X_{iK}$ to be able to estimate the average treatment effect?

Because W_i is randomly assigned, W_i and all $X_1, \ldots, X_k$ are independent. Suppose we run the regression:

$$Y_i = \beta_0 + \tau W_i + \tilde{\epsilon}_i$$

The error term $\tilde{\epsilon}_i$ incorporates the omitted X_k's:

$$\tilde{\epsilon}_i = \beta_1 X_{i1} + \cdots + \beta_k X_{iK} + \epsilon_i$$

Define $\mu_k = \mathbb{E}(X_{ik})$, and define $\delta = \sum_{k=1}^{K} \beta_k \mu_k$. Then we can write this error term as

$$\begin{aligned} \tilde{\epsilon}_i &= \delta + \sum_{k=1}^{K} \beta_k (X_{ik} - \mu_k) + \epsilon_i \\ &= \delta + \nu_i \end{aligned}$$

Because of unconfoundedness:

$$\mathbb{E}(X_{ik} - \mu_k | W_i) = \mathbb{E}(X_{ik} | W_i) - \mu_k = \mathbb{E}(X_{ik}) - \mu_k = 0.$$

Hence,

$$\begin{aligned}\mathbb{E}(\nu_i|W_i) &= \mathbb{E}\left(\sum_{k=1}^{K}\beta_k(X_{ik}-\mu_k)+\epsilon_i|W_i\right)\\ &= \sum_{k=1}^{K}\beta_k\mathbb{E}(X_{ik}-\mu_k|W_i)+\mathbb{E}(\epsilon_i|W_i)\\ &= 0\end{aligned}$$

We can therefore write the regression in this form:

$$Y_i = (\beta_0+\delta)+\tau W_i+\nu_i$$

The estimate of the intercept will be $\beta_0+\delta$, and because $\mathbb{E}(\nu_i|W_i)=0$, we will obtain an unbiased estimate of the ATE, τ.

In practice, if all or some of the X_k's are observed, we should still include them in the regression because the error variance will be reduced and the estimated coefficients will be more precise.

Potential outcomes and regression analysis

So far we considered the binary treatment case, $W_i = 0, 1$

More generally, we can think of causality in regression analysis using the potential outcome $Y(X)$,

$$Y(X) = \mathcal{C}(X)+\epsilon = \beta_0+\beta_1X_1+\cdots+\beta_kX_K+\epsilon$$

- $\mathcal{C}(X)$ is the *causal regression function*
- All variables combined, $X=(X_1,\ldots,X_K)$, are the treatment, and $Y(X)$ is the potential outcome given the treatment assignment X
- Can be generalized to any non-linear causal regression function

The causal effect of treatment assignment X' vs. X while *fixing* ϵ is given by

$$\begin{aligned} Y(X') - Y(X) &= (\mathcal{C}(X') + \epsilon) - (\mathcal{C}(X) + \epsilon) \\ &= \mathcal{C}(X') - \mathcal{C}(X) \end{aligned}$$

Example 1: The causal effect of increasing X_k to $X'_k = X_k + 1$ while leaving all other inputs unchanged is

$$\mathcal{C}(X') - \mathcal{C}(X) = \beta_k$$

Example 2: A demand curve consists of all the potential outcomes = quantities, $Q = Y(X)$, given the marketing mix X (price, promotion, advertising, ...). The demand curve predicts quantities when we change X while fixing ϵ

To understand causality in a regression model it is crucial to distinguish between the demand curve as described by the causal regression function $\mathcal{C}(X)$ from the *statistical* regression function used to fit the data,

$$Y = \tilde{\beta}_0 + \tilde{\beta}_1 X_1 + \cdots + \tilde{\beta}_K X_K + \tilde{\epsilon}$$

For any data, we can fit such a regression and obtain what is called the **best linear predictor**. The residual $\tilde{\epsilon}$ based on the best linear predictor will always satisfy:

$$\begin{aligned} \mathbb{E}(X_k \tilde{\epsilon}) &= 0 \\ \mathbb{E}(\tilde{\epsilon}) &= 0 \end{aligned}$$

The first property says that the residuals are uncorrelated with the covariates. Note that for the second property to be true we need to include an intercept in the regression.

However, if the true error terms in the potential outcome ($Y(X) = \mathcal{C}(X) + \epsilon$) do not satisfy $\mathbb{E}(\epsilon|X) \neq 0$, then then the coefficients of the best linear predictor are not the same as the coefficients in the causal regression function, $\beta_k \neq \tilde{\beta}_k$.

Causality in regression analysis: Conclusion

We can always fit a regression model to the data, and we will obtain the best linear predictor.

The estimates will not represent the causal regression function unless $\mathbb{E}(\epsilon|X) \neq 0$. We cannot learn if we obtained the causal regression function from inspecting the residuals, because the residuals will always be uncorrelated with the covariates.

We obtain the causal regression function if we manipulate the inputs, X, through randomization in a randomized controlled trial.

Causal effects in observational studies

- Observational studies
  - Treatment is not randomly assigned under *experimental conditions*
- Can we ever interpret the estimated effects in observational studies as causal?

As good as randomly assigned

Let's illustrate using a binary treatment $W_i = 0, 1$. Even if the treatment W_i is not fully random, it can be random conditional on the observed pre-treatment variables $X_{i1}, \ldots, X_{iK}$.

Correspondingly, the fully general statement of unconfoundedness is:

1. **Unconfoundedness** (general): The potential outcomes, $Y_i(0)$ and $Y_i(1)$ are statistically independent of the treatment W_i *conditional* on $X_{i1}, \ldots, X_{iK}$

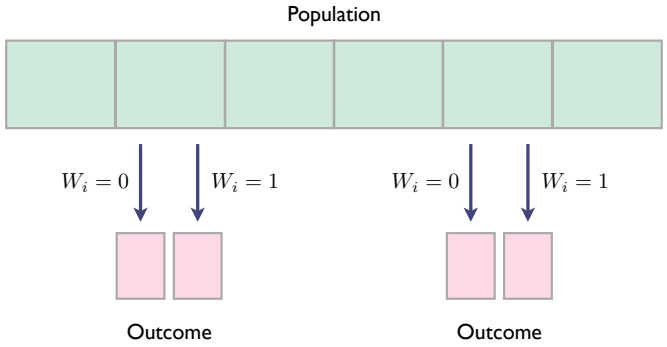
Loosely speaking, we then say that the treatment W_i is *as good as randomly assigned*. If satisfied, we can estimate the ATE (average treatment effect) based on

$$\mathbb{E}[Y_i^{\text{obs}}|W_i = 1] - \mathbb{E}[Y_i^{\text{obs}}|W_i = 0] = \text{ATE}$$

Random assignment in homogenous subpopulations

Units with identical values of the pre-treatment variables $X_{i1}, \ldots, X_{iK}$ form *homogenous subpopulations.*

The idea of general unconfoundedness is that the treatment assignment is random within the homogenous subpopulations.



Hence we can estimate the conditional treatment effect in each subpopulation, then average over all subpopulations to obtain the overall average treatment effect (ATE).

Estimating average treatment effects in observational studies

If the treatment W_i is as good as randomly assigned in the homogenous sub-populations, we can estimate the average treatment effect τ using the regression

$$y_i = \beta_0 + \tau W_i + (\beta_1 x_{i1} + \cdots + \beta_k x_{iK}) + \epsilon_i$$

- Note: Now we **must** control for the variables $x_{i1}, \ldots, x_{iK}$

In a more general regression model: W_i has multiple levels or is continuous

Examples: Estimating advertising effects

If advertising budget is set as a fixed fraction of sales: Advertising is not as good as randomly assigned across markets. But if we control for market fixed effects and if there are no temporal confounders, then advertising is as good as randomly assigned *within* each market.

Thomas (2016): Advertising for antihistamines. Advertising is not as good as randomly assigned over time because pollen levels are correlated with advertising and pollen levels influence demand. Conditional on pollen levels, however, advertising is as good as randomly assigned over time.

Shapiro (2016) border strategy: Even at local market level advertising may be correlated with demand fluctuations, and we do not have the market research information on these fluctuations that the advertiser had in the past. Solution: Estimate advertising effects based on difference in advertising levels across DMA border. Assumption: Conditional on a common time trend, the *differences* in advertising levels in the two bordering DMA's are as good as randomly assigned over time.

Natural experiments

- Natural experiments
  - Data where we have good reason to believe that the treatment is as good as randomly assigned
  - Example: Shapiro border experiments
- Natural experiments are not as powerful as randomized controlled trials to infer causal effects
  - "As good as randomly assigned" is an assumption
  - Assumption can **never** be directly tested — recall discussion of best linear predictor

Background on the potential outcomes model

- Potential outcomes model is also called the *Neyman-Rubin causal model*
  - Jerzey Neyman (1894 – 1981)
  - Donald Rubin (1943 –)
- Increasingly popular framework among statisticians, social scientists, and researchers in the biomedical sciences
- Ongoing active debate with proponents of alternative (dissenting) approaches to causality

Application: Income and wealth effects on private label demand

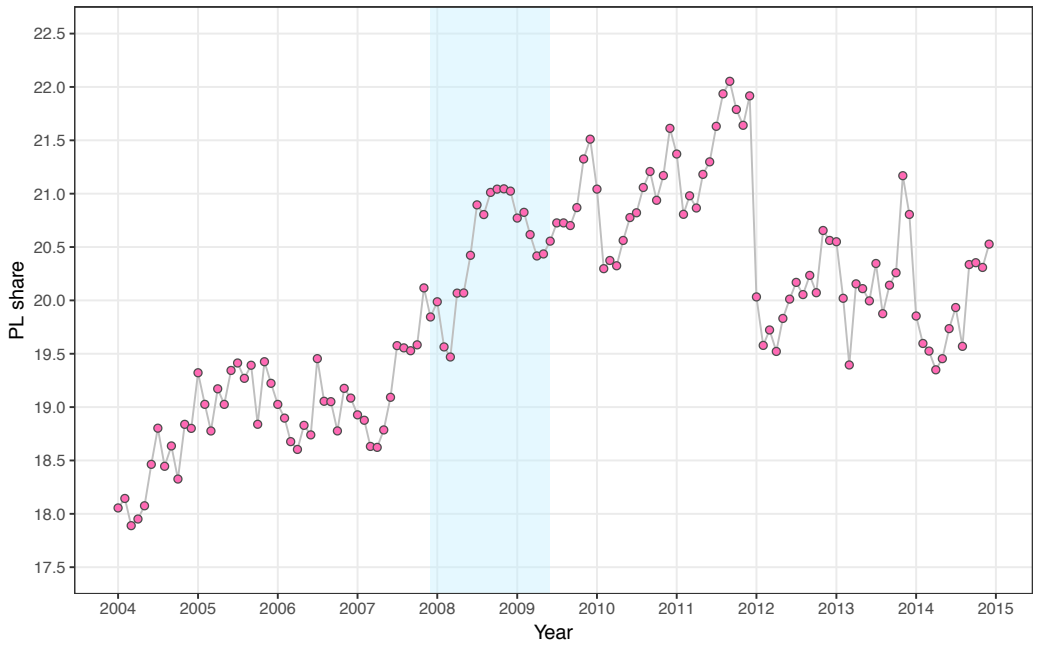
Dubé, Hitsch, and Rossi (2016): "Income and Wealth Effects on Private Label Demand: Evidence From the Great Recession"

Private label products: Important element of retail strategy, corresponding concern for national brand manufacturers

Industry concern about effects of income and wealth

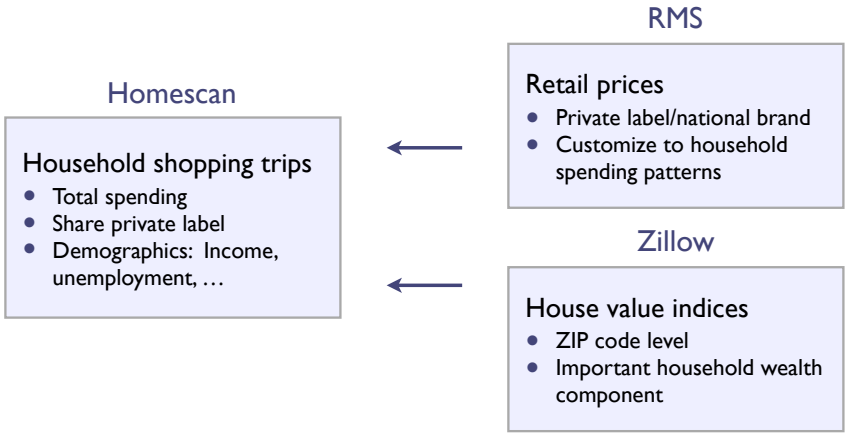
- Particular interest because of Great Recession of 2007-2009 and aftermath
- "One thing you don't want to do is create a consumer who shifted to private label and then have to spend a lot to get them back."
  Tom Falk, chairman/CEO Kimberly-Clark
- "We've done studies spanning decades, and what we document, very clearly, is that private label grows a lot in recessions."
  Jan-Benedict Steenkamp, Professor of Marketing at UNC Kenan-Flager School of Business

Evolution of private label share, 2004 - 2014



Data sources to study private label demand: Nielsen Homescan and RMS at the Kilts Center

- Extremely comprehensive data
- Match private label expenditure share and
  - Income, unemployment
  - Wealth (Zillow housing value)
  - Price levels



Causal effects of income and wealth

- Goal: Measure income/wealth effects on private label demand
- Ideal data source:
  - Randomly applied changes in income and wealth to different households
  - Feasibility
- Our data
  - Income, wealth (home value), and unemployment status observed
  - Observational study

Utilize income and wealth variation across households

- Income, wealth, and unemployment varies across households
- Estimate income and wealth effects using regression analysis
  - Dependent variable
    - Private label expenditure share
  - Independent variables
    - Income
    - Wealth (housing value)
    - Unemployment indicator
    - Demographics
- *Can we interpret the income/wealth estimates as causal?*

	(1)
log(Income)	-2.55
	(0.0967)
log(Wealth)	-1.35
	(0.146)
$\mathbb{I}\{\text{Unemployed}\}$	0.00438
	(0.13)
Controls	Demographics
No. Obs.	1,716,936
R^2	0.0723

Interpreting the income/wealth estimates

- Simple case: two income (or wealth) levels, high and low
  - Low income: $W_i = 0$
  - High income: $W_i = 1$
- The regression estimate of the income effect measures

$$\tau = \mathbb{E}[Y_i|W_i = 1] - \mathbb{E}[Y_i|W_i = 0]$$

  - Difference in private label share between high and low income households
- Is this the causal treatment effect of an income increase (from low to high)?

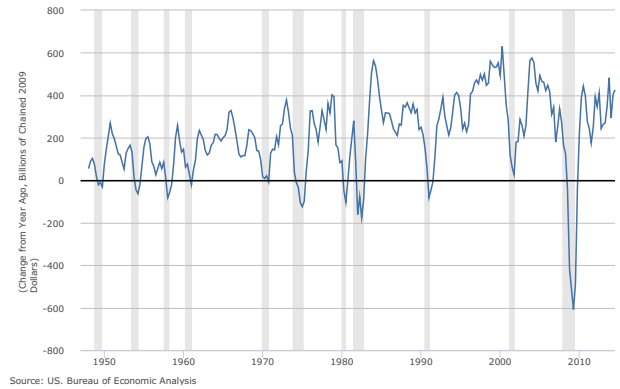
- Income is not randomly assigned to the households
  - People choose careers with different income levels
    - True even controlling for skills
  - People who choose careers with high incomes may also have stronger preferences for branded products than people in low-income professions

- Possible interpretation of the income effect estimate in the regression
  - Households with high incomes spend less on private label products (as a percentage of total expenditure) than low income households
  - May reflect difference in preferences and need not be causal

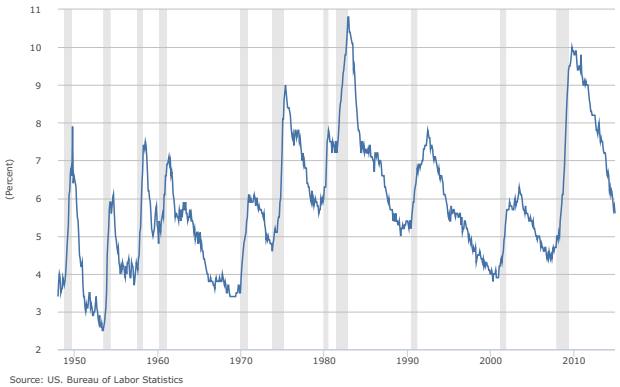
- How can we fix this problem?

Important variation in the data: The Great Recession

- Large changes in real economic activity
  - Drop in real GDP
  - Increase in unemployment



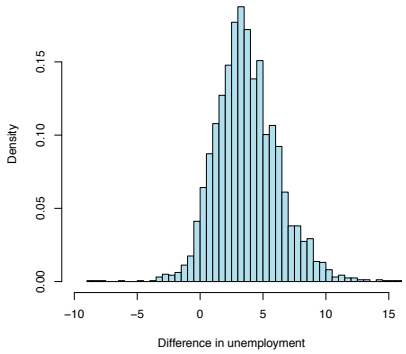
Source: US. Bureau of Economic Analysis

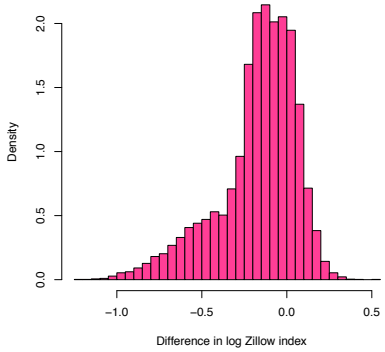


Source: US. Bureau of Labor Statistics

Recession effects across regions and households

- Compare January 2006 vs December 2009
- Large regional differences in
  - Change in unemployment
  - Change in housing values
- Large overall changes and large regional differences:
  - *Great Recession as a natural experiment*





Utilize within-household variation in income and wealth

- Control for systematic differences in private label preferences across households
  - Household fixed effects
  - Infer income and wealth effects from within-household changes in income and wealth over time
- When can we interpret the estimated effects as causal?
  - Assumption: Within-household income and wealth changes are as good as randomly assigned
  - Argue this is plausible because main source of income and wealth changes in the data is the Great Recession
    - Large and sudden impact
    - Large differences in recession impact on income and wealth across households and regions

Indirect evidence to support "as good as random assignment"

- Document the correlation between
  1. Difference in private label share between two periods t_0 and t_1 before the approximate onset of the Great Recession
  2. Difference in market-level unemployment, market-level house prices, and household income between two periods during the Great Recession
- Idea
  - Suppose we find that, for example, lower housing prices are associated with an increase in private label share during the Great Recession
  - Was there already a trend towards private label before the Great Recession in local areas with a subsequent large housing price decline?

Trend analysis: Correlation coefficients

Pre-Recession	Recession	Estimate	t-statistic	Pr(>|t|)	No. obs.
Correlation PL Share/Unemployment					
2004/2006	2007/2009	-0.042	-1.974	0.049	2,185
2005/2006	2007/2009	-0.007	-0.331	0.741	2,183
Q1 2004/Q1 2007	Q2 2007/Q2 2009	-0.027	-1.267	0.205	2,147
Q1 2005/Q1 2007	Q2 2007/Q2 2009	0.018	0.861	0.389	2,195
Q1 2006/Q1 2007	Q2 2007/Q2 2009	-0.012	-0.562	0.574	2,199
Correlation PL Share/log Zillow Index					
2004/2006	2007/2009	0.001	0.109	0.913	5,985
2005/2006	2007/2009	-0.028	-2.135	0.033	5,934
Q1 2004/Q1 2007	Q2 2007/Q2 2009	-0.001	-0.068	0.946	5,793
Q1 2005/Q1 2007	Q2 2007/Q2 2009	-0.027	-2.051	0.040	5,985
Q1 2006/Q1 2007	Q2 2007/Q2 2009	-0.022	-1.720	0.086	6,067
Correlation PL Share/log Income					
2004/2006	2007/2009	0.010	1.342	0.180	17,207
2005/2006	2007/2009	-0.005	-0.738	0.460	19,689
Q1 2004/Q1 2007	Q2 2007/Q2 2009	-0.004	-0.569	0.569	17,207
Q1 2005/Q1 2007	Q2 2007/Q2 2009	-0.014	-2.020	0.043	19,681
Q1 2006/Q1 2007	Q2 2007/Q2 2009	-0.006	-0.891	0.373	21,591

Note: Analysis is at the county (unemployment), ZIP code (Zillow), and household (income) level

Estimates using within-household variation

- Income effect much smaller when we control for household fixed effects

	Original estimates	Controlling for fixed effects
	(1)	(2)
log(Income)	-2.55	-0.479
	(0.0967)	(0.0669)
log(Wealth)	-1.35	-2.38
	(0.146)	(0.119)
$\mathbb{I}\{\text{Unemployed}\}$	0.00438	0.268
	(0.13)	(0.0837)
Controls	Demographics	Household FE
No. Obs.	1,716,936	1,716,947
R^2	0.0723	0.567

Controlling for trends, level shifts, and prices

- (3): Add controls for
  - Time trend
  - Trend interaction during/after recession
  - Private label share level shift during/after recession
- (4): Add controls for
  - Private label/national brand prices
  - Overall price level

	(1)	(2)	(3)	(4)
log(Income)	-2.55	-0.479	-0.412	-0.38
	(0.0967)	(0.0669)	(0.0654)	(0.0713)
log(Wealth)	-1.35	-2.38	-0.545	-0.16
	(0.146)	(0.119)	(0.141)	(0.164)
$\mathbb{I}\{\text{Unemployed}\}$	0.00438	0.268	0.373	0.353
	(0.13)	(0.0837)	(0.0812)	(0.0906)
Trend			0.0372	0.0368
			(0.00159)	(0.00456)
Trend $\times \mathbb{I}\{\text{Recession}\}$			0.019	0.011
			(0.00419)	(0.00655)
Trend $\times \mathbb{I}\{\text{Post Rec.}\}$			-0.029	-0.0295
			(0.0039)	(0.00578)
$\mathbb{I}\{\text{Recession}\}$			0.322	0.138
			(0.0427)	(0.0489)
$\mathbb{I}\{\text{Post Recession}\}$			0.479	0.24
			(0.0769)	(0.111)
Controls	Demographics	Household FE	Household FE	Household FE Prices
No. Obs.	1,716,936	1,716,947	1,716,947	1,317,596
R^2	0.0723	0.567	0.57	0.578

Controlling for income lags

- (5): Add income lags
  - Past income (1 - 4 years)
- Results
  - Evidence that permanent income changes have larger effect than current income changes

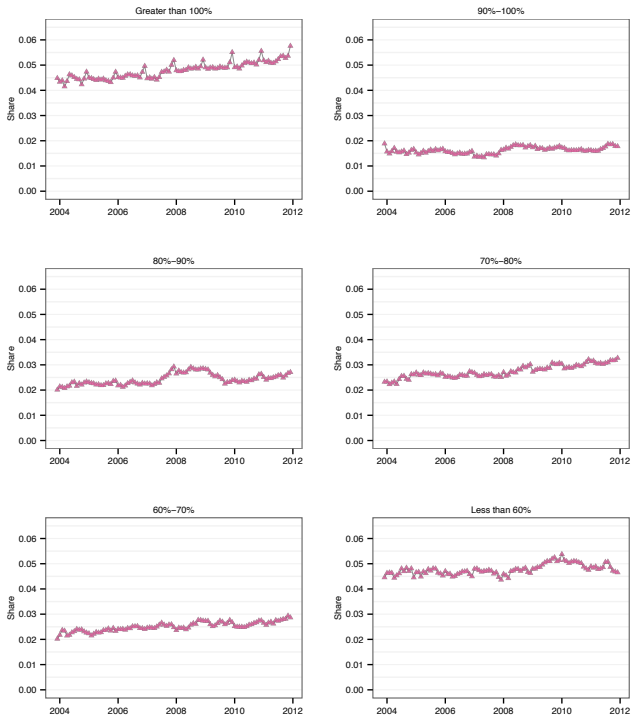
	(4)	(5)
log(Income)	-0.38	-0.282
	(0.0713)	(0.0839)
log(Wealth)	-0.16	0.0269
	(0.164)	(0.198)
$\mathbb{I}\{\text{Unemployed}\}$	0.353	0.372
	(0.0906)	(0.113)
$\log(\text{Income})_{-1}$		-0.438
		(0.088)
$\log(\text{Income})_{-2}$		-0.292
		(0.0853)
$\log(\text{Income})_{-3}$		-0.0984
		(0.0853)
$\log(\text{Income})_{-4}$		-0.22
		(0.0844)
Controls	Household FE Prices Trend/Level shifts	Household FE Prices Trend/Level shifts
No. Obs.	1,317,596	899,423
R^2	0.578	0.581

Interpreting the effect sizes

- 25 percent income decline
  - Contemporaneous effect: 0.11 percentage point increase in private label share
  - Effect from permanent income decline: 0.38 point increase
- Effect of trend on private label share
  - 0.44 percentage point increase per year
  - Very large compared to income effects
- Income effect magnitudes
  - Much smaller than claimed by industry insiders
  - Much smaller compared to published results in marketing research
    - Example: Lamey et al. (2007): "How Business Cycles Contribute to Private- Label Success: Evidence from the United States and Europe," *Journal of Marketing Research*, 71(1)
    - Claims 1.22 percent permanent increase in private label share growth for a 1 percent decline in per-capita GDP

Sources of trend in private label share?

- Private label share trend not systematically related to quality tier
- Quality tier: Based on relative private label/national brand price



- New product introduction rate not different across private label and national brand products

Conclusions

- Great Recession as a natural experiment
  - Large changes in income and wealth
  - Large regional differences in income and wealth changes

- Causal negative effect of income on private label share
  - Robust to household fixed effects, trends, level shifts during/after recession, and prices
  - "As good as randomly assigned" is plausible given the controls, hence effect likely to be *causal*
  - Permanent income change has larger effect than contemporaneous income change
  - Effect sizes: small!

- Large overall increase in private label share in U.S. due to trend
  - Sources of trend
    - Not concentrated in specific quality tiers (e.g. growth stemming from high quality private label products)
    - Not related to relatively more private label compared to national brand product introductions

Summary

- Potential outcomes model ("parallel worlds")
  - Causality: Difference across potential outcomes

- Measuring causality — average treatment effect
  - Randomized controlled trials (A/B testing)
    - Unconfoundedness (satisfied because of random assignment) implies causality
  - Observational studies and natural experiments
    - Unconfoundedness (treatment is as good as randomly assigned) implies causality
  - RCT's always more powerful than observational studies, because in observational studies unconfoundedness cannot be directly tested

- Application: Measuring income and wealth effects on private label demand
  - Natural experiment to infer causal effects