Lecture Note 17: Private Information, Adverse Selection and Market Failure

14.03/14.003, Microeconomic Theory and Public Policy, Fall 2010
David Autor, Massachusetts Institute of Technology

December 1, 2010

1 Private Information, Adverse Selection and Market Failure

- Where there is private information, there is an incentive for agents to engage in strategic behavior. For example, if you are selling a product, and your buyer knows the distribution of product quality but not the quality of the individual product that you posses, how much should the buyer be willing to pay? The intuitive answer might be the *expected value* of the product, or perhaps the certainty equivalent of this lottery.

- But this answer ignores the an important consideration: The choice of what product you sell may depend on what price the buyer offers. And the price that the buyer offers may depend on what product she thinks you'll sell at that price. The equilibrium outcome in which buyer and seller expectations are aligned—that is, the buyer gets what she wants at the price she offers—may be far from efficient.

1.1 A bit of background

- Economists had historically conjectured that markets for information were well-behaved, just like markets for other goods and services. One could optimally decide how much information to buy, and hence equate the marginal returns to information purchases with the marginal returns to all other goods.

- In the 1970s, economists were brought to reevaluate this belief by a series of seminal papers by Akerlof, Rothschild-Stiglitz, and Spence. These economists all went on to share the 2001 Nobel for their work on the economics of information. We will discuss

all three contributions. (The coauthor on the Stiglitz paper, Michael Rothschild, did not receive the Nobel.)

- Information is not a standard market good:

  - Non-rivalrous (no marginal cost to each person knowing it)

  - Extremely durable (not consumed)

  - Not a typical experience good where you can 'try before you buy.' Cannot readily allow you to 'sample' information without actually giving you information.

  - Unlike other goods (or their attributes), information is extremely difficult to measure, observe, verify.

- This combination of odd properties often gives rise to settings where information is—at least potentially—*asymmetric*. That is, some agents in a market are better informed than others about the attributes of a product or transaction.

- The most natural (and surely ubiquitous) way in which this occurs is that buyers may have general information about the 'average' characteristics of a product that they wish to purchase whereas sellers will have *specific* information about the individual product that they are selling.

- When buyers and sellers have asymmetric information about market transactions, the trades that are transacted are likely to be a subset of the feasible, welfare-improving trades. Many potentially Pareto-improving trades will not be completed due to informational asymmetries (that is, trades that would voluntarily occur *if all parties had full information* will not take place).

- Economic models of information are often about the information environment—who knows what when. Specifying these features carefully in the model is critical to understanding what follows.

- These notes discuss two key results:

  1. The "Lemons Principle"
  2. The "Full Disclosure Principle"

- It turns out that these principles are roughly inverses.

2 Adverse Selection: The Market for Lemons (Akerlof, 1970)

2.1 The fundamental problem:

1. Goods of different quality exist in the marketplace.

2. Owners/sellers of goods know more about their goods' quality than do buyers.

3. Critical insight of Akerlof: *Potential buyers know that sellers know more about the quality of goods than they do.*

- This information asymmetry can substantially affect the market equilibrium. It is possible that there will be *no trade* whatsoever for a given good, although:

    1. At any given price p_0, there are traders willing to sell their products.

    2. At price p_0, there are buyers willing to pay strictly above p_0 for the good that traders wold like to sell.

- George Akerlof (1970) was the first economist to analyze this paradox rigorously. His paper was nominally about the market for used cars. It's always been folk wisdom that it's a bad idea to buy used cars—that 'you are buying someone else's problem.' But why should this be true? If used cars are just like new cars only a few years older, why should someone else's used car be any more problematic than *your* new car after it ages a few years?

2.2 A simple example: The market for used cars

- Setting

    - There are 2 types of *new* cars available at dealerships: good cars and lemons (which break down often).

    - The fraction of lemons at a dealership is λ.

    - Dealers do not distinguish (perhaps by law) between good cars versus lemons; they sell what's on the lot at the sticker price.

    - Buyers cannot tell apart good cars and lemons. But they know that some fraction $\lambda \in [0, 1]$ of cars are lemons.

- After buyers have owned the car for any period of time, they also can tell whether or not they have bought a lemon.

  - Assume that good cars are worth $V_g = \$2,000$ to buyers and lemons are worth $V_l = \$1,000$ to buyers.

  - For simplicity (and without loss of generality), assume that cars do not deteriorate and that buyers are risk neutral.

- What is the equilibrium price for *new* cars? This will be

$$P_n = (1 - \lambda) \cdot 2,000 + \lambda \cdot 1,000.$$

- Since dealers sell all cars at the same price, buyers are willing to pay the *expected value* of a new car.

- Now, consider the used car market. Assume that used car sellers are willing to part with their cars at 20 percent below their new value. So,

$$S_g^u = \$1,600 \text{ and } S_l^u = \$800.$$

- Since cars don't deteriorate, used car buyers will be willing to pay $\$2,000$ and $\$1,000$ respectively for used good cars and lemons. There is a surplus of $\$400$ or $\$200$ (that is, a gain from trade) from each sale. Selling either a good car or a lemon is potentially Pareto improving.

- Question: What will be the equilibrium price of used cars?

- The intuitive answer is

$$P_u = E\left[V | Used\right] = (1 - \lambda) \cdot 1,600 + \lambda \cdot 800.$$

But this is not necessarily correct.

- Recall that buyers cannot distinguish good cars from lemons, while owners of used cars know which is which. Assuming (logically) that sellers will only part with their cars if offered a price that is greater than or equal to their reservation price. So, for $P_u \geq 800$, owners of lemons will gladly sell their cars. However, for $P_u < 1,600$, owners of good cars will *keep* their cars.

- Denote buyers' willingness to pay for a used car as B_u.

- If there will be trade in equilibrium, buyers' willingness to pay must satisfy the following inequality: $B_u\left(E\left[S_u\left(P_u\right)\right]\right) \geq P_u$. That is, at price P_u, the quality of cars available for sale, $S_u\left(P_u\right)$, must be worth at least that price to buyers.

- Given the reservation selling prices of $S_{u,l}, S_{u,g}$ the quality of cars available depends on the price. In particular, the share of Lemons is as follows:

$$\Pr\left(\text{Lemon}|P_u\right) = \begin{cases} 1 & \text{if } P_u < 1,600 \\ \lambda & \text{if } P_u \geq 1,600 \end{cases}.$$

That is, quality is *endogenous* to price. More specifically:

$$E\left[S_u\left(P_u\right)\right] = \begin{cases} 800 & \text{if } P_u < 1,600 \\ 800 \cdot \lambda + (1-\lambda) \cdot 1,600 & \text{if } P_u \geq 1,600 \end{cases}$$

- What is $B_u\left(E\left[S_u\left(P_u\right)\right]\right)$? The value to buyers of cars for sale as a function of price is:

$$B_u\left(P_u\right) = \begin{cases} 1000 & \text{if } P_u < 1,600 \\ \lambda \cdot 1000 + (1-\lambda) \cdot 2000 & \text{if } P_u \geq 1,600 \end{cases}$$

- The willingness of buyer's to pay for used cars depends upon the market price (a result we have not previously seen in consumer theory).

- Take the case where $\lambda = 0.4$. Consider the price $P_u = 1,600$. At this price, the expected value (to a buyer) of a randomly chosen used car—assuming both good cars and lemons are sold—would be

$$B_u\left(P_u = 1,600, \lambda = 0.4\right) = (1 - 0.4) \cdot 2000 + 0.4 \cdot 1000 = 1,600.$$

Here, used *good* cars sell at exactly the average price at which potential sellers value them. Owners of good cars are indifferent and owners of lemons get a $800 surplus. This equation therefore satisfies the condition that $B_u\left(S_u\left(P_u\right)\right) \geq P_u$.

- But now take the case where $\lambda = 0.5$. At price $P_u = 1,600$, the expected value of a randomly chosen used car is:

$$B_u\left(P_u = 1,600, \lambda = 0.5\right) = (1 - .5) \cdot 2000 + .5 \cdot 1000 = 1,500.$$

- $B_u\left(S_u\left(P_u\right)\right) < P_u$. This cannot be an equilibrium. Since owners of good used cars demand $1,600$, they will not sell their cars at $1,500$. Yet, $P_u = 1,500$ is the maximum price that buyers will be willing to pay for a used car, given that half of all cars are lemons. Consequently, good used cars will *not* be sold in equilibrium, despite the fact that they are worth more to buyers than to sellers. Thus, *only* lemons sell.

- More generally, if $\lambda > 0.4$, then good used cars are not sold and $P_u \in [800, 1000]$. In this price range, $B_u E\left[S_u\left(P_u\right)\right] \geq P_u$.

- Bottom line: If the share of lemons in the overall car population is high enough, the bad products drive out the good ones. Although buyers would be willing to pay $2,000$ for a good used car, their inability to distinguish good cars from lemons means that they are not willing to pay more than $1,500$ for *any* used car. With λ high enough, no good cars are sold, and the equilibrium price must fall to exclusively reflect the value of lemons.

2.3 Summing up the Akerlof adverse selection model

- The key insight of Akerlof's paper is that market quality is *endogenous,* it depends on price. When sellers have private information about products' intrinsic worth, they will only bring *good* products to market when prices are high.

- Buyers understand this, and so must adjust the price they are willing to pay to reflect the quality of the goods they expect to buy at that price.

- *In equilibrium, goods available at a given price must be worth that price.* If they are not, then there will be no equilibrium price and it's possible that no trade will occur (which is the case in the lemons model in the Akerlof paper).

3 Equilibrium in models with asymmetric information

One feature that makes models with asymmetric information somewhat different from models we've studied previously is that equilibrium is not primarily determined by a set of marginal conditions. e.g. marginal profit is zero, or the marginal rate of substitution equals the price ratio. Instead, equilibrium is a strategic notion. We think of parties on the different sides of the market (e.g., buyers v. sellers) as choosing strategies (feasible actions) that maximize their payoffs given the chosen strategies of the players on the other side of the market. But of course, the players on the other side of the market are likewise choosing strategies to maximize

their payoffs given the actions (or anticipated actions) of the other players. An equilibrium in this setting is a set of complementary strategies such that neither side wants to unilaterally change its strategy given the strategy of the other side. This notion is what is called a Nash Equilibrium after John Forbes Nash, who developed the idea and proved its existence in a 28 page 1950 Princeton doctoral dissertation, which eventually won him the Nobel prize in Economics in 1994.

Here's an informal definition (paraphrased from Wikipedia): A Nash Equilibrium is a solution concept for a game involving two or more players, in which each player is assumed to know the equilibrium strategies of the other players, and no player has anything to gain by changing only his or her own strategy unilaterally. If each player has chosen a strategy and no player can benefit by changing his or her strategy while the other players keep theirs unchanged, then the current set of strategy choices and the corresponding payoffs constitute a Nash equilibrium.

We saw this idea in the used car example above:

- We first worked out the strategies of used car sellers, taking as given the price offered by buyers. We concluded that if the offer price was less than $1,600$, only lemons were sold, whereas if the offer price was $\geq 1,600$, both lemons and good cars were sold.

- We observed that buyers have two primary strategies available: offering 800 and offering 1600.

- We then asked which of these buyer strategies constituted a Nash equilibrium given the strategies of sellers.

- Clearly, offering 800 is always a Nash equilibrium. If a buyer offers 800, sellers will offer only lemons. Since these cars are worth 800, the buyer and seller strategies constitute a Nash equilibrium. Neither party wishes to deviate from their strategy (e.g., offer more, sell a good car) given the strategy of the other player.

- We next asked whether offering $1,600$ could also be a Nash equilibrium. The answer, as we saw, depends upon λ, the population share of lemons. At offer price $1,600$, both lemons and good cars are sold (that's the sellers' strategy). For this to be a Nash equilibrium, buyers must be happy to pay a price of $1,600$ given the sellers' strategies when facing this price. We calculated that the value to buyers of cars available at offer price $1,600$ is:

$$E\left[B_u | P_u = 1,600\right] = (1 - \lambda) \cdot 2000 + \lambda \cdot 1000.$$

For $P_u = 1,600$ to be a Nash equilibrium, it must be the case that $E\left[B_u\left(P_u = 1,600\right)\right] \geq 1,600$, which requires that $\lambda \leq 0.4$

We have been implicitly (and explicitly) using the notion of Nash equilibrium to solve the asymmetric information models that we've considered so far, including Rothschild-Stiglitz and Akerlof.

4   Adverse selection: A richer example

- Now that we have seen a stylized example, let's go through the same logic with a slightly richer example. We will consider a continuous distribution of product quality rather than just two types.

- Consider the market for 'fine' art. Imagine that sellers value paintings at between $0 and $100,000, denoted as V_s, and these values are uniformly distributed, so the average painting is worth $50,000 to a seller.

- Assume that buyers value paintings at 50% above the seller's price. Denote this valuation as V_b. If a painting has $V_s = \$1,000$ then $V_b = \$1,500$.

- The only way to know the value of a painting is to buy it and have it appraised. Buyers cannot tell masterpieces from junk. Sellers can.

- What is the equilibrium price of paintings in this market?

- An equilibrium price must satisfy the condition that the goods that sellers are willing to sell at this price are worth that price to buyers: $V_b\left(E\left[V_s\left(P\right)\right]\right) \geq P$.

- Take the sellers' side first. A seller will sell a painting if $P \geq V_s$.

- There is a range of sellers, each of whom will put their painting on the market if $P \geq V_s$.

- What is the *expected* seller's value of paintings for sale as a function of P? Given that paintings are distributed uniformly, it is:

$$E\left[V_s | P\right] = \frac{0 + P}{2}.$$

So, if $P = 100,000$ then *all* paintings are available for sale and their expected value to sellers is $50,000. If $P = 50,000$, the expected seller value of paintings for sale is $25,000.

- Now take the buyer's side. Since the $V_b = 1.5 \cdot V_s$, buyers' willingness to pay for paintings as a function of their price is

$$E\left[V_b | E\left[V_s | P\right]\right] = 1.5 \cdot E\left[V_s | P\right] = 1.5\left(\frac{0 + P}{2}\right) = \frac{3}{4}P.$$

Clearly $E\left[V_b | E\left[V_s | P\right]\right] < P$. No trade occurs

- Since that buyers' valuation of paintings lies strictly above sellers' valuations, this outcome is economically inefficient—that is, the gains from trade are unrealized. What's wrong?

- The sellers of low-quality goods generate a negative externality for sellers of high quality goods. For every \$1.00 the price rises, seller value only increases by \$0.50 because additional low-quality sellers crowd into the market $\left(\frac{\partial E[V_s|P_s]}{\partial P} = 0.5\right)$.

- Consequently, for every dollar that the price rises, buyers' valuations only increases by \$0.75, $\left(\frac{\partial E[V_b|E[V_s|P]]}{\partial P} = 0.75\right)$. There is no equilibrium point where the market price 'calls forth' a set of products that buyers are willing to buy at that price.

- This is in effect the "Lemons Principle"—The goods available at a given price are worth less than or equal to that price (to sellers).

- In this example, there is no trade.

5 Reversing the Lemons equilibrium: The Full Disclosure Principle

- Is there a way around this result? Intuition would suggest that the answer is yes. Sellers of *good* products have an incentive to demonstrate the quality of their products so that they can sell them at their true value. (Sellers of bad products have an incentive to not disclose quality, and this is what 'spoils' the market.)

- In the example above, sellers of good products do not disclose their products' quality because we have stipulated that the value of a piece of art can only be assessed *ex-post* by appraisal. Sellers of good paintings therefore have no credible means to convey their products' quality.

- Needed: A means to disclose information credibly. If there is an inexpensive (or free) means to credibly disclose the quality of paintings, you might expect that sellers of *above*

average quality paintings will probably want to do this. In actuality, the result is much stronger than this: all sellers will choose to disclose.

5.1   Simplest case: Costless verification

- Imagine now that a seller of a painting can get a free appraisal. This appraisal will credibly convey the true seller's value of the painting (and so the buyer's willingness to pay will be 1.5 times this value). Who will choose to get their paintings appraised?

- Your first instinct might be that, since buyers are willing to pay $75,000$ for a painting of average quality, any seller with a painting that would sell for at least $75,000$ if appraised would choose to get an appraisal.

- This intuition is on the right track but incomplete. It neglects the fact that the decision by some sellers to have their paintings appraised affects buyers' willingness to pay for non-appraised paintings.

- If only sellers with $V_s \geq 75,000$ had their paintings appraised, what would be the market price of non-appraised paintings?

$$1.5 \cdot E\left[V_s | V_s < 75,000\right] = 56,250.$$

- But if the market price is only $56,250$, then sellers with paintings at or above this price will also get them appraised. What is the new market price of non-appraised paintings?

$$1.5 \cdot E\left[V_s | V_s < 56,250\right] = 42,888.$$

- And so on...

- You can keep working through this example until you eventually conclude that *all* sellers will wish to have their paintings appraised. Why? Because each successive seller who has his painting appraised devalues the paintings of those who do not. This in turn causes additional sellers to wish to have their paintings appraised. In the limit, the only seller who doesn't have an incentive to obtain an appraisal is the seller with $V_s = 0$. This seller is indifferent.

- This example demonstrates the *Full-Disclosure Principle*. Roughly stated: If there is a credible means for an individual to disclose that he is above the average of a group, she

will do so. This disclosure will implicitly reveal that other non-disclosers are below the average, which will give them the incentive to disclose, and so on... If disclosure is costless, in equilibrium all parties will explicitly or implicitly disclose their private information. If there is a cost to disclosure, there will typically be a subset of sellers who do not find it worthwhile to disclose.

- The Full Disclosure Principle is essentially the inverse of the Lemons Principle. In the Lemons case, the bad products drive down the price of the good ones. In the Full Disclosure case, the good products drive down the price of the bad ones. What distinguishes these cases is simply whether or not there is a credible disclosure mechanism (and what the costs of disclosure are).

5.2   Costly verification

- Imagine now that a seller of a painting must pay $5,000$ for a appraisal. Which paintings will be appraised? If there are non-appraised paintings, will they be sold and at what price?

- We now need to consider three factors simultaneously:

  1. The net price of a painting that the seller would obtain if the painting were appraised (net of the appraisal fee)

  2. The net price if not appraised

  3. The value of the painting to the seller (remember that sellers won't sell for a net price less than V_s).

- The following conditions must be satisfied in equilibrium:

  1. Buyer's willingness to pay for an appraised painting is greater than or equal to seller's value of painting:
  $$V_b\left(A=1\right) \geq V_s + 5000,$$
  We will refer to this as Individual Rationality Constraint 1 ($IR1$).

  2. Buyer's willingness to pay for a non-appraised painting is greater than or equal to seller's value of painting:
  $$V_b\left(A=0\right) \geq V_s$$
  We will refer to this as Individual Rationality Constraint 2 ($IR2$).

3. Seller cannot do better by appraising a non-appraised painting or v.v. We will refer to this as the Self-Selection Constraint (SS). Consider a cutoff value V_s^*. In equilibrium Paintings with $V_s \geq V_s^*$ are appraised and paintings with $V_s < V_s^*$ are not:

$$V_b\left(V_s \geq V_s^*, A = 1\right) - 5,000 \geq V_b\left(V_s \geq V_s^*, A = 0\right)$$

and

$$V_b\left(V_s < V_s^*, A = 1\right) - 5,000 \leq V_b\left(V_s < V_s^*, A = 0\right).$$

- Let's go through these one at a time.

1. Rewriting $IR1$, we have:

$$\begin{aligned} V_b\left(A = 1\right) &\geq V_s + 5000, \\ 1.5V_s &\geq V_s + 5000. \end{aligned}$$

2. Rewriting $IR2$, we have:

$$\begin{aligned} V_b\left(A = 0\right) &\geq V_s, \\ 1.5 \cdot E\left[V_s \leq V_s^*\right] &\geq V_s, \\ 1.5 \cdot E\left[V_s \leq V_s^*\right] &\geq V_s^*. \end{aligned}$$

3. Rewriting SS, to solve for the critical value of V_s^*:

$$\begin{aligned} V_b\left(V_s = V_s^*, A = 1\right) &\geq V_b\left(V_s = V_s^*, A = 0\right), \\ 1.5V_s^* - 5,000 &\geq \frac{1.5V_s^*}{2}. \end{aligned}$$

(This implicitly satisfies the second inequality in SS as well: $V_b\left(V_s < V_s^*, A = 1\right) \leq V_b\left(V_s < V_s^*, A = 0\right)$.)

- Let's solve these out of order.

1. Solving $IR1$: $1.5V_s \geq V_s + 5000 \Rightarrow V_s \geq 10,000$. That is, no painting under $\$10,000$ would be appraised because the purchase price at the appraised value would not compensate the seller for his reservation price.

2. Solving $IR2 : 1.5 \cdot E\left(V_s \leq V_s^*\right) \geq V_s^* \Rightarrow \frac{3}{4}V_s^* \geq V_s^* \Rightarrow V_s^* = 0$. $IR2$ can only be satisfied with $V_s^* = 0$. That is, as in the pure adverse selection case above, non-appraised paintings *cannot* be sold. There is no market in non-appraised paintings because buyers' willingness to pay for them is lower than sellers' valuation of these paintings.

3. Solving SS for V_s^* gives $\frac{3}{4}V_s^* \geq 5,000 \Rightarrow V_s^* = \$6,666$.

- Combining these results, we have

$$
\begin{aligned}
V_s^* &= 10,000, \\
P\left(A = 1\right) &\geq \max\left[V_s^* + 5,000, 1.5 \cdot V_s^*\right], \\
P\left(A = 0\right) &= 0,
\end{aligned}
$$

That is, paintings with $V_s \geq 10,000$ are appraised and sold at $\leq 1.5 \cdot V_s^*$ but with a minimum price of $15,000$ (the lowest price that a seller of an appraised painting with $V_s = 10,000$ would accept). Note that $IR1$ and SS implied different values of the cutoff value of V_s^* ($10,000$ and $6,666$ respectively). Only $IR1$ binds because sellers' valuation of paintings is sufficiently high that they are unwilling to accept the market price for an non-appraised painting. Thus, the operative constraint is not that the *market price* for an non-appraised painting is higher than the *market price* for an appraised painting (which is SS) but that the *seller's own valuation* of an appraised painting net of appraisal cost must be greater than the *market price* of that painting when appraised (which is $IR!$).

- Paintings that are not appraised are not sold because buyers would be willing to pay no more than $\$7,500$ for them if all were sold. But at that price, only paintings worth up to $\$7,500$ would be sold, meaning that buyers would only be willing to pay $\$5,625$, and so on...

- Conclusion from this example: If verification of information is costly, the market equilibrium will not be Pareto efficient. In particular, there are two distortions evident in this market equilibrium. First, most sellers are spending $\$5,000$ to appraise and sell their paintings, even though this investment does nothing to improve the painting (so, this is a deadweight loss). Second, paintings with $V_s < \$10,000$ are not sold, even though these paintings are worth $1.5 \cdot V_s$ to buyers.

6   Summary

- Unobservable quality heterogeneity creates important problems for market efficiency—market failures or incomplete markets quite likely.

- The problem is not the uncertainty *per se*. As we demonstrated during the lectures on risk, uncertainty and the market for insurance, there are market mechanisms for trading efficiently in risk. In these models, the uncertainty is *exogenous*—it is not under the control of economic agents.

- The fundamental problem in the adverse selection models above is that asymmetric information leads to a market equilibrium where sellers use their informational advantage strategically. Buyers respond strategically to sellers' choices. And the equilibrium of these strategic games are not likely to be first-best efficient. Quality is endogenous to price.

6.1   Market responses to asymmetric information

- If Lemons (adverse selection) hypothesis is correct, there should be some market mechanisms *already in place* to ameliorate the problem. Conversely, if no economic agent was observed attempting to solve the problem, we would have reason to doubt that the Lemons problem is relevant.

- What are some of these mechanisms?

  - Private mechanisms: Information provision, warranties, brand names, specialists and testers.

  - Licensing.

  - Mandated information provision.

  - Legal liability.

  - Regulation.

  - Example: Health insurance 'open enrollment' periods. Life insurance applications.

  - Lemon laws.

- Are there any markets that simply don't exist because of adverse selection (or moral hazard)?

14

- Lifetime income insurance

- Why is health insurance so expensive for the unemployed?

- Why doesn't my life insurance policy cover suicide during the first five years after purchase?

- Why can't I buy insurance against getting a low GPA at MIT?

7 Analyzing adverse selection: The case of subprime loans

This section refers to the 2010 paper in the *Quarterly Journal of Economics* by Keys, Mukherjee, Seru, and Vig. I will discuss the substantive details of the paper in class. In these notes, I will discuss the framework for causal inference.

Referring to our potential outcomes framework, imagine that for each individual i, there there exists a pair of potential outcomes: Y_{i1} for what would occur if the unit were exposed to the treatment and Y_{i0} if not exposed. The causal effect of the treatment is represented by the difference $T = Y_{i1} - Y_{i0}$. As usual, the fundamental problem of causal inference is that we cannot observe the pair Y_{i1} and Y_{i0}. We have typically handled this problem through difference-in-difference estimation, sometimes using an explicit experimental randomization to assign units to treatment versus control conditions. We have also used quasi-experiments (such as the minimum wage increase in NJ) and instrumental variables (such as Air-Sea Differential Distance). All of these methods attempt to find units that are in expectation comparable—that is, their potential outcomes if treated (or if untreated) do not differ in expectation—and then compare outcomes among those that are treated relative to those who are not treated.

The Regression Discontinuity (RD) approach takes a novel to identifying a causal relationship when the treatment and control groups do not have potential outcomes that are identical in expectation. It instead looks for units that are *arbitrarily close* in terms of their potential outcomes and yet are treated differently (one assigned to treatment, the other assigned to control) due to some bright line rule used to assign them. This situation occurs more commonly than one might expect. For example, the result of a national election can be decided by a single vote, or the cutoff for which children are allowed to enter 1st grade in a given year may depend on whether they were before or after midnight on September 1 six years earlier.

Why are such arbitrary cutoffs useful? Define a variable X that is used to determine the cutoff. For example, X could be the percentage of voters for candidate A or X could be the

exact hour of birth. Now, imagine there are two underlying relationships between potential outcomes and X, represented by $E[Y_{i1}|X_i]$ and $E[Y_{i0}|X_i]$. And let's say that individuals to the right of some cutoff c (let's say $X_i \geq 0.5$) are exposed to treatment, and all those to the left ($X_i < 0.5$) are denied treatment. Therefore, we only observe $E[Y_{i1}|X_i]$ to the right of the cutoff and $E[Y_{i0}|X_i]$ to the left of the cutoff.

As we consider units i that are arbitrarily close to the threshold, it would be reasonable to assume that:

$$\lim_{\varepsilon \downarrow 0} E[Y_{i1}|X_i = c + \varepsilon] = \lim_{\varepsilon \uparrow 0} E[Y_{i1}|X_i = c + \varepsilon],$$
$$\lim_{\varepsilon \downarrow 0} E[Y_{i0}|X_i = c + \varepsilon] = \lim_{\varepsilon \uparrow 0} E[Y_{i0}|X_i = c + \varepsilon].$$

That is, for units that are *almost identical*, we can loosely say that if they had both been treated (or not treated), their outcomes would have been near-identical. If so, then we can form a Regression Discontinuity estimate of the causal effect of treatment on outcome Y using the contrast:

$$\hat{T} = \lim_{\varepsilon \downarrow 0} E[Y_i|X_i = c + \varepsilon] - \lim_{\varepsilon \uparrow 0} E[Y_i|X_i = c + \varepsilon],$$

which in the limit is equal to:

$$T = E[Y_{i1} - Y_{i0}|X_i = c].$$

This is the key idea underlying the RD technique used for causal inference by Keys et al.