Lecture Note 2 – Causal Inference in Economics, with an Application to the Minimum Wage Debate

David Autor, MIT Department of Economics

14.03/14.003 Microeconomic Theory and Public Policy, Fall 2016

1 Causal inference in Economics

In addition to studying this lecture note, please be sure to give a careful read to Chapter 1 in Mastering 'Metrics "Randomized Experiments," which does a very nice job with this material. For those who'd like a deeper dive, I encourage you to also read the 1986 JASA article "Statistics and Causal Inference" by Paul Holland. Causal inference will be at the foundation of 14.03/14.003, and we will build on this foundation throughout the semester.

Much of social science (psychology, anthropology, sociology, political science, epidemiology and some subset of economics) concerns analyzing correlations among variables, e.g., the correlation between education and income, the correlation between obesity and heart disease, the correlation between happiness and longevity.

- Correlation describes the statistical relationship between two observed variables. Correlation has no necessary relationship with cause and effect. You can measure the correlation between happiness and longevity with great precision and yet know nothing about whether making someone happier will improve their longevity. (Let's say that happier people live longer. Perhaps these people are happy precisely because they feel fortunate to have lived so long.)

- In 14.03/14.003, we are not deeply interested in these correlations. Science advances primarily through analyzing cause and effect relationships, not by documenting correlations (though correlations are not irrelevant).

- Causal questions:

  - What is the causal effect of education on income?
  - What is the causal effect of obesity on heart disease?
  - What is the causal effect of happiness on longevity?
  - What is the causal effect of the minimum wage on employment?

- Causal questions are much harder to answer than the correlational questions. Here's why:

  - *Correlations are readily measured from observational data.*
  - *Causal effects can never be measured directly (to be explained).*

- The reason that causal questions are difficult to answer is that they intrinsically concern a *counterfactual* state—that is, something that by definition has not or did not occur.

- To be concrete, let's examine a simple case where we have a binary causal variable $X \in \{0, 1\}$ and a response variable Y (which may be continuous).

  - If X has a causal effect on Y, this implies that the value of Y we observe depends on the value of X.

- We can define Y_0 and Y_1 as the values that Y would take if X were equal to 0 and 1 respectively.
- Y_0 and Y_1 are counterfactuals of one another. If $X = 0$ then we will have $Y = Y_0$. The counterfactual question is what value of Y would we have observed if instead we had $X = 1$. Obviously, the answer is Y_1. But Y_1 is a notional concept. We can posit that both Y_0 and Y_1 have well defined values, but we will never see both of them.

1.1 The fundamental problem of causal inference (FPCI)

- Let Y_i be the outcome of interest for unit i, where i could be a person, a cell, a drop of water, a sovereign country. We'll suppress the i subscript where possible.
- We want to consider two possible outcomes for i. Let Y_0 be the value of Y_i for $X = 0$ and Y_1 be the value of Y_i for $X = 1$, as above. Thus, for every unit i, we can imagine two potential outcomes $\{Y_0, Y_1\}$ that we would observe if the unit were untreated ($X = 0$) or treated ($X = 1$).
- For the moment, let's assume that people are all alike, so $Y_{0i} \equiv Y_0$ and $Y_{1i} \equiv Y_1$.
- We only observe either Y_0 or Y_1, but we assume that both are well defined. That is, there is a precise alternative state of the world that would have occurred had we chosen $X = 1$ instead of $X = 0$ or vice versa.
- In this framework, the causal effect of X_i on Y_i is $T = Y_1 - Y_0$, where T stands for Treatment Effect.
- The problem that this immediately reveals is that *we never observe* $Y_1 - Y_0$ for a single unit i. Instead, we observe

$$Y_i = Y_{1i}X_i + Y_{0i}(1 - X_i).$$

  That is, we observe Y_1 or Y_0 but not both.

Definition 1. Fundamental Problem of Causal Inference. It is not possible to observe the value Y_{1i} and Y_{0i} for the same unit i, so we cannot measure the causal effect of X on Y for unit i.

- Natural question: Why can't we just switch X from 0 to 1 and back to observe both Y_1 and Y_0? In fact, this procedure is not informative about $Y_1 - Y_0$ without further assumptions (discussed below).
- One useful observation to build intuition: many causal relationships are irreversible. If X corresponds to attending MIT vs. another college and Y corresponds to post-college earnings, we can either observe your post-MIT earnings or your post-non-MIT earnings, not both.

2 Addressing the fundamental problem of causal inference

Since the problem is fundamental, there is no solution. But there are several "work-arounds."

2.1 Work-around 1: Postulate stability and reversibility (AKA 'causal transience')

- One work around is to assume stability and reversibility (what Paul Holland calls *temporal stability* and *causal transience*).
- If the causal effect of X on Y is the same at every point in time (now and the future) *and* the causal effect of X on Y is reversible (so having once been exposed X doesn't permanently change the effect of X on Y), then we can observe $Y_{1i} - Y_{0i}$ simply by repeatedly changing X from 0 to 1.
- Formally, these assumptions are: $Y_{1it} = Y_{1i}$, $Y_{0it} = Y_{0i}$ for all i and t where t indexes time.
- Note that temporal stability and causal transience are *postulates*. They cannot be tested
- Example: You can turn water from ice to steam and back repeatedly to analyze the causal effect of temperature change on water molecules. But what allows you to make the *causal inference* that steam is the counterfactual for ice when the treatment is 100 degrees versus 0 degrees Celsius are the postulates that (1) water molecules are not fundamentally altered by heating and cooling; and (2) that the relationship between temperature and the behavior of water is stable (e.g., does not depend on the phase of the moon).
- Counter-example: It would probably not be valid to assess the effectiveness of a treatment for high cholesterol for patient i by repeatedly administering the cholesterol reducing treatment, testing the patient's cholesterol level, then withdrawing the treatment, testing the patient's cholesterol level, etc. Cholesterol levels are sluggish state variables. And they might be permanently affected by even a one-time treatment.

2.2 Work-around 2: Postulate homogeneity

- We may alternatively assume *unit homogeneity*. If Y_{1i} and Y_{0i} are identical for all i, we can measure the causal effect of X on Y simply by taking the difference $Y_{1i} - Y_{0j}$ for $i \neq j$.
- Of course, unit homogeneity is also a postulate; one cannot know that two things are identical in all respects.
- But under certain laboratory conditions, unit homogeneity seems quite reasonable (e.g., experimenting with two molecules of water). This assumption would clearly be invalid for two cholesterol patients. Or for any two people more generally.

2.3 Work-around 3: Estimate causal effects for populations rather than individuals

- For human subjects, neither (1) temporal stability and causal transience nor (2) unit homogeneity can plausibly be expected to hold in any setting. No two people are alike. And no one

person is identical to him or herself at a different point in time.

- *We should therefore acknowledge that we will never be able to credibly estimate* $T_i = Y_{1i} - Y_{0i}$ for a person i.

- We might, however, be satisfied to settle for some kind of population average treatment effect instead:

$$T^* = E\left[Y_1 - Y_0 | X = 1\right],$$

  where $E[\cdot]$ is the *expectations* operator, denoting the mean of a random variable. This expression above defines the Average Treatment Effect for the Treated (ATT), that is the causal effect of the treatment on the people who received the treatment (i.e., for whom $X = 1$).

- The ATT should be distinguished from the Average Treatment Effect (ATE), defined as

$$T^\dagger = E\left[Y_1 - Y_0\right].$$

  The difference between T^* and $T^\dagger$ is this: ATT only measures the causal effect for those who receive treatment whereas the ATE is the causal effect one would notionally obtain if *everyone* were treated. In practice, these can be quite different. The ATT for a cholesterol lowering drug given to morbidly obese patients is probably not comparable to the ATE for a cholesterol lowering drug given to the entire population of adults. We are usually more interested in knowing the ATT than the ATE.

- Returning to our discussion of ATT, how do estimate this quantity. Since we cannot directly observe T for any given individual i, how do we measure $E[Y_1 - Y_0|X = 1]$ for some population of $i's$?

- One idea: We could compare $E[Y|X = 1]$ and $E[Y|X = 0]$ to form $\tilde{T} = E[Y|X = 1] - E[Y|X = 0]$. For example, let X be the cholesterol treatment and Y be a measure of serum cholesterol level. We could compare cholesterol levels among those taking the treatment ($E[Y|X = 1]$) versus those not taking the treatment ($E[Y|X = 0]$) to estimate the causal effect of the treatment on cholesterol levels. Is this a good idea?

- A moment's thought should suggest that $\tilde{T}$ is *not* a good estimator for T^*. The problem is that people who take the cholesterol treatment are likely to have abnormally high cholesterol whereas those who do not take the treatment are likely to have normal cholesterol levels. Thus, *even if* the treatment lowered cholesterol, we might erroneously conclude the opposite because our comparison group ($X = 0$) had low cholesterol initially whereas our treatment group ($X = 1$) had abnormally high cholesterol—and may still have above average cholesterol, even if the treatment is somewhat effective.

- So, if $\tilde{T}$ is not a good measure of T^*, what would a good comparison look like? We need to find treatment and control populations that have the same expected levels of cholesterol *but for* the

treatment. Formally, we want to identify a set of people for whom the *counterfactual* outcomes are comparable between the treatment and comparison (AKA control) groups. Specifically:

$$\begin{aligned} E\left[Y_1|X=1\right] &= E\left[Y_1|X=0\right] \\ E\left[Y_0|X=1\right] &= E\left[Y_0|X=0\right]. \end{aligned} \tag{1}$$

- If these conditions are satisfied. then it's straightforward to see that a contrast of the outcomes of the treatment and control groups will provide a valid estimate of the causal effect of treatment for the treated group. Specifically,

$$\begin{aligned} E\left[Y_1|X=1\right] - E\left[Y_0|X=0\right] &= E\left[Y_1|X=1\right] - E\left[Y_0|X=1\right] \\ &= T^* \end{aligned}$$

  Notice that our substitution above of $E\left[Y_0|X=1\right]$ for $E\left[Y_0|X=0\right]$ is justified by the assumption of treatment-control balance in equation (1). If the subjects who didn't receive the treatment are just like those who did *but for* not having received the treatment, then the contrast between the treated and untreated groups provides an unbiased estimate of the causal effect of the treatment on the treated group (ATT).

- Returning to our *invalid* estimator $\tilde{T}$, let's ask how likely is it that the counterfactual outcomes would be balanced among a set of people selected from the population according to whether or not they are currently receiving the treatment. It does not take a great leap of logic to hypothesize that the patients receiving the cholesterol drug are more likely to be suffering from high cholesterol than those who are not taking the drug. This would imply that:

$$\begin{aligned} E\left[Y_1|X=1\right] &> E\left[Y_1|X=0\right] \\ E\left[Y_0|X=1\right] &> E\left[Y_0|X=0\right]. \end{aligned}$$

  In words, patients receiving the drug are more likely to suffer from high cholesterol *whether or not* they are receiving the drug (presumably, they are receiving the drug specifically because they were diagnosed with high cholesterol).

- So, if we calculated the contrast $\tilde{T} = E\left[Y|X=1\right] - E\left[Y|X=0\right]$ for this unbalanced group, what would we get?

$$\begin{aligned} E\left[Y_1|X=1\right] - E\left[Y_0|X=0\right] = &\underbrace{E\left[Y_1|X=1\right] - E\left[Y_0|X=1\right]}_{T^*} \\ &+ \underbrace{\left\{E\left[Y_0|X=1\right] - E\left[Y_0|X=0\right]\right\}}_{Bias}. \end{aligned} \tag{2}$$

  The first term on the right-hand side of this equation is the true, causal effect of the cholesterol treatment on those who take it (the ATT). The second term is the potential bias that

occurs if the counterfactual (non-treated) outcomes of the comparison group (those not taking the treatment) differ from the counterfactual (non-treated) outcomes of the treatment group (those taking treatment) if they were untreated. [Note that the expression above simplifies to $E[Y_1|X=1] - E[Y_0|X=1]$. The substitutions cancel out.]

- We've just argued above that $E[Y_0|X=1] > E[Y_0|X=0]$. Thus, the bias in this case is positive—that is, it goes in the direction of generating an estimate of $\tilde{T}$ that is *larger* than the true casual effect T^* (so, $E\left[\tilde{T}\right] > E[T^*]$). Even if T^* were hypothetically negative (that is, the drug reduces cholesterol), we could easily conclude through this naive comparison that the drug increases cholesterol levels.

2.4 Another example

- Let's say Y is the number of mathematical expressions you can differentiate in an hour after 4 years of college and X is an indicator variable for whether or not you attended MIT.

- If we administered math tests at random to students at Boston area colleges, we would certainly find that $\tilde{T} = E[Y_1|MIT=1] - E[Y_0|MIT=0] > 0$, i.e., MIT students can solve more calculus problems in an hour than non-MIT students.

- But $\tilde{T}$ a not a valid estimate of the causal effect of attending MIT on calculus skills (that is, $\tilde{T} \neq E[Y_1 - Y_0|\text{MIT}=1]$). Students who are skilled in calculus choose MIT, and they would be more skilled than the average student in calculus, regardless of whether they attended MIT. So, $E[Y_0|MIT=1] > E[Y_0|MIT=0]$.

- The substantive problem (again) is that the "treatment," MIT attendance, is endogenous. Students come to MIT in part because they are good at math. It is unwise to assume that non-MIT students are a valid comparison group for MIT students.

3 Implementing the statistical solution using randomization

- The most robust, time-tested, and intuitive solution to the FPCI for causal inference with human subjects (or complex organisms) is random assignment.

Definition 2. Random assignment is an experimental technique for assigning human participants or animal subjects to different groups in an experiment (e.g., a treatment group versus a control group) using randomization, such as by a chance procedure (e.g., flipping a coin) or a random number generator. This ensures that each participant or subject has an equal chance of being placed in any group. Random assignment of participants helps to ensure that any differences between and within the groups are not systematic at the outset of the experiment. Thus, any differences between groups recorded at the end of the experiment can be more confidently attributed to the experimental procedures or treatment. [Source Wikipedia, accessed 9/4/2016.]

- Let's say that we picked a large number of $i's$ at random and randomly assigned half to $MIT = 1$ and half to $MIT = 0$.

- With a large sample, we are pretty much guaranteed that

$$E[Y_1|MIT = 1] = E[Y_1|MIT = 0] \text{ and } E[Y_0|MIT = 1] = E[Y_0|MIT = 0].$$

  So, condition (1) should be satisfied.

- Plugging back into (2):

$$\hat{T} = E[Y_1|MIT = 1] - E[Y_0|MIT = 0] =$$
$$E[Y_1|MIT = 1] - E[Y_0|MIT = 1] + \underbrace{\{E[Y_0|MIT = 1] - E(Y_0|MIT = 0)\}}_{bias\ =0}.$$

- Thus, randomization has eliminated the bias term (in expectation) by balancing the counterfactual outcomes between the treatment and control groups. Specifically, the students assigned to MIT would have been expected to fare comparably to the students who were *not* assigned to MIT had these students instead been assigned to MIT. What randomization has bought us specifically is this:

$$\{E[Y_0|MIT = 1] - E[Y_0|MIT = 0]\} = 0.$$

- In summary, randomization overcomes the causal inference problem by making the treatment status $X = \{0, 1\}$ independent of potential outcomes: $E(Y_1), E(Y_0) \perp X$ so

$$E[Y_1|X = 1] = E[Y_1|X = 0] \quad \text{and} \quad E[Y_0|X = 1] = E[Y_0|X = 0]$$

  Another way of saying this is that randomization makes the treatment and control groups *exchangeable*, so we can be confident that the differences in outcomes we observe between them are caused by their assignment to treatment vs. control status rather than underlying differences between the two populations.

- Randomly assignment subjects to the treatment and controls ensures that the group not receiving the treatment provides a valid estimate of the *counterfactual* outcome for the treated group.

- Bottom line:

  - It is rarely plausible for human behavior that either of the two alternative solutions will be plausible (temporal stability + causal transience or unit homogeneity). By contrast, so long as we can randomize, the statistical solution is likely to work.
  - Needless to say, to solve the Fundamental Problem of Causal Inference in economics, we always use the statistical solution. That is, we never assume that humans are homo-

geneous or temporally stable. In some cases, we use randomized experiments. Where randomized experiments are infeasible or impractical, we use quasi-experiments.

4 Difference-in-Difference Estimation

- In many experimental settings, we not only compare treatment and control subjects after the treatment has been administered (at *endline*) but also prior to treatment (at *baseline*). Collecting pre and post-treatment data allows us to measure the change in the outcome variable of interest for each subject and to compare the average change in the outcome variable for the treatment versus control populations.

- Putting this into practice, if we have a treatment and control group, we can form:

	Before	After	Change
Treatment	Y_{j0}	Y_{j1}	ΔY_j
Control	Y_{k0}	Y_{k1}	ΔY_k

- Why do we want to make a pre-post comparison?

- We actually do not need to do this if we have a very large population of (randomly assigned) treatment and control units to work with. In that case, we could simply calculate

$$\hat{T} = E\left[Y|X=1\right] - E\left[Y|X=0\right] = E\left[Y_1 - Y_0|X=1\right].$$

  If X is randomly assigned and the population of treated units is large, then the conditions in equation (1) should apply and hence the cross-sectional (as opposed to over-time, or inter-temporal) comparison should provide a valid estimate of the causal effect of interest.

- However, we often don't have very large samples of treatment and control individuals to work with.

- Let's say we are assessing the effect of a new drug treatment on cholesterol levels. We could pick 10 people each for the treatment and control groups, give the treatment group the drug treatment and the control group the placebo, and then compare the average cholesterol level between these two groups.

- There is nothing wrong with this approach. But we might be concerned that, just by chance, these two groups started out with somewhat different cholesterol levels.

- Because of this concern, we could also take baseline data (prior to treatment) to ensure that these groups were comparable.

- Let's say the baseline averages were comparable but not identical; by chance, the treatment group had a slightly lower cholesterol level than the treatment group. We'd be concerned that our experiment would be biased in favor of the finding that the treatment lowered cholesterol (since the treatment group started with a better outcome).

- It's that concern that motivates us to compare the *change* in cholesterol in the treatment group to the change in cholesterol in the control group. By studying the change in the outcome variable, we subtract off initial differences in levels that could potentially prove confounding in small samples. Thus, we focus on the *improvement* (or change) in the treatment group *relative to* the control group.

- Formally, let's say that prior to treatment:

$$Y_{j0} = \alpha_j.$$
$$Y_{k0} = \alpha_k.$$

  We would hope that $\alpha_j \simeq \alpha_k$, but this does not strictly have to be the case.

- Now, imagine that after treatment, we observe

$$Y_{j1} = \alpha_j + \delta + T,$$

  where T is the causal effect and δ is any effect of time. For example, cholesterol levels may tend to rise over time as people age.

- So, if we take the first difference for Y_j, we get:

$$\Delta Y_j = Y_{j1} - Y_{j0} = (\alpha_j - \alpha_j) + \delta_j + T$$

  This does not recover T. But it does remove the "level effect" α_j.

- Similarly, $\Delta Y_k = (\alpha_k - \alpha_k) + \delta_k$. Differencing removes the level effect for group j.

- If we are willing to postulate that the time effect operates identically on the treatment and control groups, $\delta_j = \delta_k = \delta$, then we have

$$\Delta Y_j - \Delta Y_k = T + \delta - \delta = T.$$

- So, the difference-in-difference estimator allows us to potentially recover the causal effect of treatment even when the treatment and control groups are not entirely identical and when there is a potentially confounding effect of time.

- Summing up, there are two mean reasons for making pre-post comparisons between the treatment and control groups (that is, applying the difference-in-difference estimator):

  1. The pre-post comparison increases precision. For an outcome that tends to be relatively stable over time such as cholesterol level or calculus ability, the within-person change can be very informative. Given a choice between a sample size of 200 and only endline data or a sample size of 100 with baseline and endline data, we would often prefer the latter

because the over-time, within-person variation will have a higher signal-to-noise ratio than the larger endline-only sample. (These claims can be formalized, though I won't do that here.)

2. The pre-post comparison *differences out* initial discrepancies. By chance, the treatment and control groups might start out at initially different levels (an issue in small samples and, frequently, in quasi-experiments). In these cases, the pre-post comparison allows you to *difference* out these initial discrepancies under the *parallel trends* assumption that the time effect is identical for the treatment and control groups.

5 Back to Jersey

- Let Y_{n0} and Y_{n1} be the level of employment in New Jersey before ($t = 0$) or after ($t = 1$) the introduction of the minimum wage. The minimum wage hike occurs in 1993, so we will be comparing outcomes in 1992 to those in 1993/94.
- If we want to estimate the causal effect of the minimum wage hike on New Jersey employment, we could calculate:

$$\hat{T} = Y_{n1} - Y_{n0},$$

  which is simply the before/after change in New Jersey employment.
- What are the potential weaknesses of this estimate of the causal effect? One is that it requires assumption of temporal stability: were it not for the minimum wage hike, New Jersey employment would have remained unchanged.
- Is this plausible? Probably not. In our previous example, $Y_{j1} - Y_{j0} = T + \delta$. Our causal estimate so far conflates the true treatment effect with any incidental, contemporaneously occurring "time effects," such as a change in fast food demand in New Jersey.
- So, what do we need to improve on this experiment? We could select a group of states at random and assign the minimum wage increase to half of them and not to the other half. Then, we could compare employment in each group of states. A problem here is that this experiment is not available to us. But it's a good idea!
- Another possibility is to select a single state that we think is closely comparable to NJ and use it as our "control group." Here, that state is Pennsylvania.
- In this case, we could take baseline data in both states and then compare the change in NJ to the change in PA. This is our difference-in-difference estimator.

5.1 Card and Krueger (1994)

- The 1994 Card and Krueger article is a widely cited study of the impact of the minimum wage on employment levels. It created huge controversy in both policy circles and among academic

economists, and arguably caused millions of workers to get a legislated raise from the Clinton administration in 1995.

- April 1, 1992: the New Jersey minimum wage rose from \$4.25 to \$5.05 per hour (a sizable increase)
- Eastern Pennsylvania (bordering NJ) didn't raise the minimum wage. It maintained the Federal minimum wage of \$4.25 per hour.
- Card & Krueger surveyed 410 fast food restaurants.
- For purposes of the analysis, the pre-period is Feb-Mar 1992, and the post-period is period is Nov-Dec 1992.
- The setup:

	Before	After	Δ
NJ	Y_{n0}	Y_{n1}	ΔY_n
PA	Y_{p0}	Y_{p1}	ΔY_p

$$\hat{T} = \Delta Y_n - \Delta Y_p$$

- Table 3 in the paper shows "Per store employment"

	Before	After	Δ
NJ	20.44	21.03	$\Delta Y_n = +0.59$
PA	23.33	21.37	$\Delta Y_p = -2.16$

- $\hat{T} = 0.59 - (-2.16) = 2.76$ with a standard error of 1.36 (so, it is statistically significant at the 5 percent since the t-ratio is ≈ 2.0).
- The paper contains many more tests, but this is the basic result: $2.76 \approx 13.5\%$ increase in employment in NJ relative to PA.

5.1.1 Interpretations?

1. Monopsony

Other interpretations?

2. Hungry teens
3. Motivational effects
4. Confounding variables (shocks to PA that are not accounted for in the test)

5. Wrong venue (why did they study fast food?)

We will have much more to say in class about the interpretation of the Card and Krueger findings.

6 Brief discussion: Methodology of Economics – *or* Why Bother with Economic Theory?

Stepping back, I want to provide a brief overview of the methodology of economic analysis.

Definition 3. *Positive Economics*

(1) The study of "what is." A descriptive endeavor free of value judgements; (2) Building models to make sense of, and generalize, the phenomena we observe; (3) Making predictions based on those models.

Definition 4. *Normative Economics*

Assessing "what ought to be done." Making economic policy prescriptions. Note: Sometimes positive economics gives us all the tools we need to say that one policy is preferable to another. For example, when one policy is Pareto superior to another. (Not too many of these)

Definition 5. *Pareto Improvement*

A choice/policy/outcome that can make at least one person better off without making anyone else worse off represents a potential Pareto improvement. In general, economics says that you should look for Pareto improvements and implement them where possible. The Pareto criterion is morally quite timid (though it's not entirely uncontroversial—see the work of eminent twentieth century philosopher John Rawls).

- In reality, Pareto improving policy options are very rare. We tend to expect that people would already have made those types of improvements without any policy interventions!
- Most policy choices involve value judgements, ethical preferences, trade-offs among competing goals (e.g., employment and inflation; equity and efficiency).
- *Economic theory rarely tells you what policies to choose. But it often makes the trade-offs clear.*

6.1 Strength of economic approach to social science

- *Rigorous*: assumptions are stated, methods are formal, conclusions are *internally* consistent.
- *Cohesive:* built on a foundation of first principles and theory.
- *Refutable*: makes strong, testable (refutable) predictions, many (far from all!) of which appear correct.
- *Practical*: will help you to better understand how the world works.

6.2 Weaknesses of the economic approach

- "Economics is marked by a startling crudeness in the way it thinks about individuals and their motivations..."— Paul Krugman
- Strong, simplifying assumptions that are often unpalatable and cannot be completely right (e.g., people act rationally to pursue self-interested—distinct from selfish—objectives...)

6.3 But there is some strength in this weakness

- We have a model of the world; it is called "the world"—and it's generally too complicated to analyze in its totality, considering all factors at once.
- Economic theory typically presents a very simplified, highly stylized depiction of the world. But this can be quite helpful.
- "The test of the validity of a model is the accuracy of its predictions about real economic phenomena, *not* the realism of its assumptions"—Milton Friedman
- "A hypothesis is important if it explains much by little"—Milton Friedman
- Our approach: simple models, significant insights.

6.4 Three significant insights of economic approach

1. Economics is about "people doing the best with what they have." This observation gets you a long way in understanding human activity, both positively and normatively. You can understand much by starting from the premise that people are *trying* to make the best choices for themselves. Many alternative assumption—for example, people are largely irrational and guided by forces they do not perceive or comprehend—appear much less attractive.
2. Equilibrium—The market 'aggregates' individual choices to produce collective outcomes that are sometimes *spectacularly different* from individual decisions. (An example from 14.01: if the firms in a fully competitive marketplace are each trying to maximize profits, none makes economic profits in equilibrium, yet the market equilibrium maximizes the sum of consumer and producer surplus.)
3. Properties of equilibrium can be evaluated using the criterion of efficiency:
    - Given: Individuals are trying to make best choices for themselves. Does market equilibrium produce an outcome that cannot be improved upon without making at least one person worse off (i.e., is it Pareto efficient)?
    - There is no obvious reason to assume that markets would lead to desirable outcomes, i.e., that we couldn't do *much* better by engaging in central planning than in relying on the haphazard result of everyone selfishly making independent choices.

- Yet, one of the stunning insights of economics is that under some key conditions, the market will produce Pareto efficient outcomes. (Not to say that all Pareto efficient outcomes are desirable outcomes. But a Pareto efficient allocation is often a good starting point!)
- And, where the market does not produce Pareto efficient outcomes, theory provides insight into why this occurs, and may provide guidance on how to get to a better outcome.
- The question of when market outcomes are first-best efficient comes from the study of General Equilibrium. We will build towards this topic from the start of the class and develop it rigorously towards mid-semester.

14.03 / 14.003 Microeconomic Theory and Public Policy
Fall 2016