Research methodology and effective writing

Lecture IV - **Statistical analysis**

Matej Lorko
matej.lorko@euba.sk
www.lorko.sk

Suggested reading:

- Dudenhefer, P. (2009). A guide to writing in Economics. EcoTeach Center and Department of Economics, Duke University.
- Neugeboren, R. H., & Jacobson, M. (2005). Writing Economics. Harvard University.
- Johnson, J. B., Reynolds, H. T., & Mycoff, J. D. (2015). Political science research methods. Cq Press.
- Friedman, S., Friedman, D., & Sunder, S. (1994). Experimental methods: A primer for economists. Cambridge University Press.

Hypothesis testing

- Statements called statistical hypotheses are key to hypothesis testing. There are two types: null hypotheses and research or alternative hypotheses.

- Null hypotheses have two important characteristics: They are succinct and precise assertions about population parameters, such as a mean equals a certain value, a pair of proportions does not differ, or a numerical indicator of a relationship between two variables is zero. In many research reports, the null hypothesis (H0) is that something (for example, a mean or a proportion) equals zero. Hence, the word null-because zero represents no effect, such as no difference.

- In addition to stating a null hypothesis, researchers state another hypothesis called the research or alternative hypothesis. Researchers usually hope that they will be able to reject the null hypothesis in favor of their research hypothesis.

- In hypothesis testing-that is, making a decision about a null hypothesis-two kinds of mistakes are possible. The first type of mistake, is to reject a true null hypothesis. Statisticians call this mistake a type I error. nother possible mistake is failing to reject a null hypothesis that is false. This type of error is called a type II error.

Hypothesis testing

- The term level of statistical significance is used to refer to the probability of making a type I error. The three most common levels of statistical significance in political science are .05, .01, and .001

- Recall that if we take many samples to obtain estimates of a population parameter, our estimates will be normally distributed and cluster around the true value of the population parameter. Sampling distributions tell us the probability that our estimates fall within certain distances of the population parameter. This probability is known as the confidence level. The confidence interval refers to the range of likely values associated with a given probability or confidence level. Thus, for every confidence level, a particular confidence interval exists.

- The general form of the confidence interval is as follows: Estimated parameter value ± standard error x critical value.

TABLE 12-1 Types of Inferential Errors

	In the "Real" World, the Null Hypothesis Is . . .	
Decision is to	True	False
Accept H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

Correlation vs. Causation

- Fundamental issue faced by those doing empirical work in economics: disentangling causality from correlation.
- We say that two economic variables are correlated if they move together.
- But this relationship is causal only if one of the variables is causing the movement in the other.
- identification problem: given that two series are correlated, how do you identify whether one series is causing another?

Correlation vs. Causation

- Analysis
- Step 1: Document the correlation, that is whether data on two measures move together.
- Step 2: Assess whether the movements in one measure are causing the movements in the other.
    - For any correlation between two variables A and B, there are three possible explanations, one or more of which could result in the correlation:
        - A is causing B.
        - B is causing A.
        - Some third factor is causing both.
    - Well designed research thus needs to make valid causal inferences. Ideally, such a design does three things:
    - 1. Covariation: demonstrates that the alleged cause (call it X) does in fact covary (corelate) with the supposed effect, Y.
    - 2. Time order: The research must show that the cause preceded the effect: X must come before Y in time. After all, can an effect appear before its cause?
    - 3. Elimination of possible alternative causes, sometimes termed "confounding factors”: The research must be conducted in such a way that all possible joint causes of X and Y have been eliminated.

Golden standard for assessing causation: Randomized trials

- How can researchers address the problem of assessing causation? The best solution is through the gold standard of testing for causality: randomized trials.

- randomized trial: The ideal type of experiment designed to test causality, whereby a group of individuals is randomly divided into a treatment group, which receives the treatment of interest, and a control group, which does not.

- Randomized trials involve taking a group of volunteers and randomly assigning them to either a **treatment** group, which gets the treatment, or a **control** group, which does not. Effectively, **volunteers are assigned** to treatment or control **by the flip of a coin**.

- treatment group: The set of individuals who are subject to an intervention being studied.

- control group: The set of individuals comparable to the treatment group who are not subject to the intervention being studied.

Golden standard for assessing causation: Randomized trials

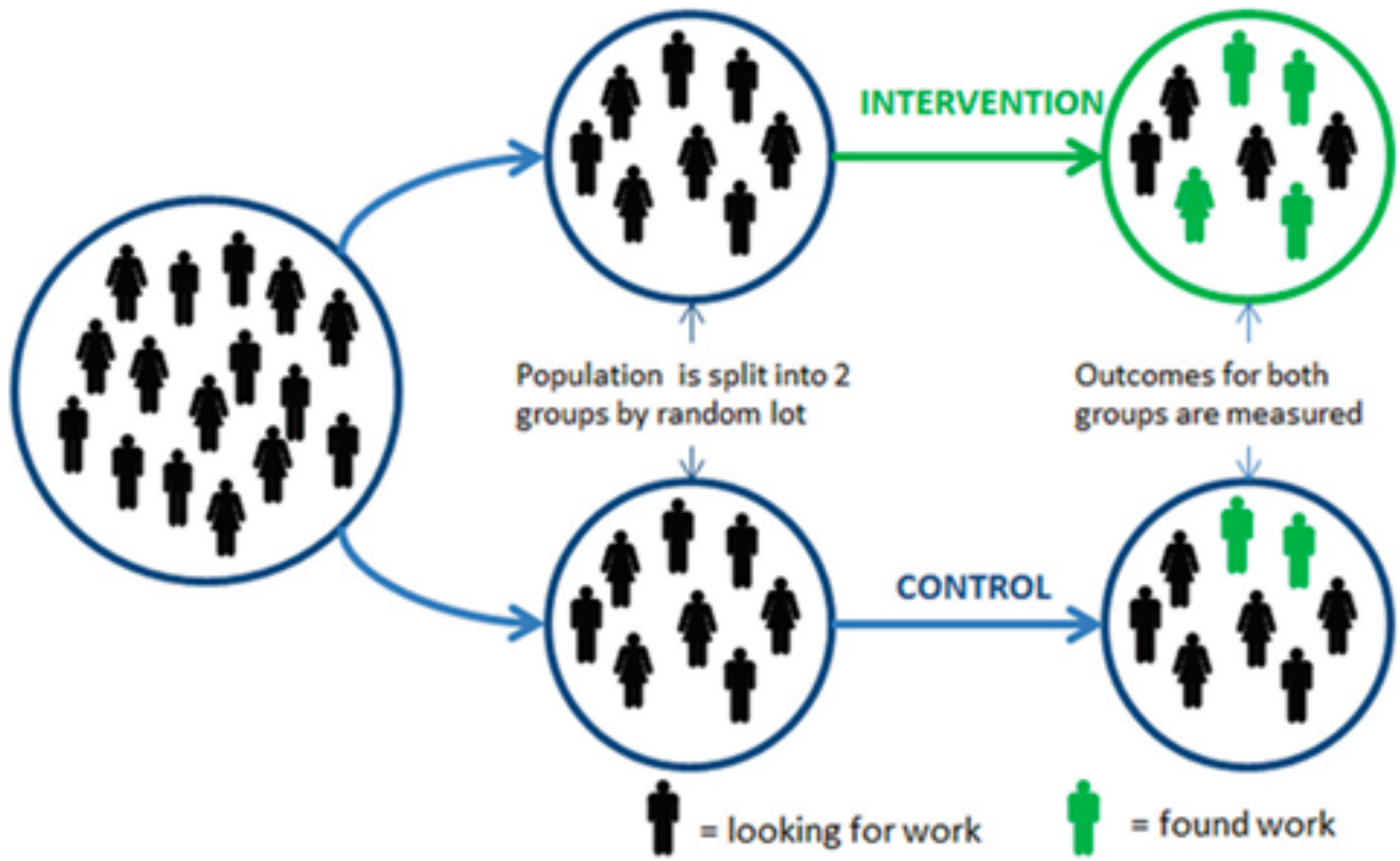


Figure 1. The basic design of a randomised controlled trial (RCT), illustrated with a test of a new 'back to work' programme.

The problem of bias

- We should always start our analysis of an empirical methodology with a simple question: Do the treatment and control groups differ for any reason other than the treatment?

- The non-treatment-related differences between treatment and control groups are the fundamental problem in assigning causal interpretations to correlations. We call these differences **bias**, a term that represents any source of difference between treatment and control groups that is correlated with the treatment but is not due to the treatment.

- By definition, such differences do not exist in a randomized trial, since the groups do not differ in any consistent fashion, but rather only by the flip of a coin.

- Thus, randomized treatment and control groups cannot have consistent differences that are correlated with treatment, since there are no consistent differences across the groups other than the treatment. As a result, randomized trials have no bias, and it is for this reason that randomized trials are the gold standard for empirically estimating causal effects.

- The description of randomized trials here relies on those trials having fairly large numbers of treatments and controls (large sample sizes). Having large sample sizes allows researchers to eliminate any consistent differences between the groups by relying on the statistical principle called the **law of large numbers**: the odds of getting the wrong answer approaches zero as the sample size grows.

Possible problems with randomized trials

- For many questions of interest, randomized trials are unfortunately not available, because they can be enormously expensive, take a very long time to plan and execute, and often raise difficult ethical issues (e.g new medical procedures)

- Moreover, even the gold standard of randomized trials has some potential problems. First, the results are only valid for the sample of individuals who volunteer to be either treatments or controls, and this sample may be different from the population at large. For example, those in a randomized trial sample may be less averse to risk or they may be more desperately ill. Thus, the answer we obtain from a randomized trial, while correct for this sample, may not be valid for the average person in the population.

- A second problem with randomized trials is that of **attrition**: individuals may leave the experiment before it is complete. This is not a problem if individuals leave randomly, since the sample will remain random. Suppose, however, that the experiment has positive effects on half the treatment group and negative effects on the other half, and that as a result the half with negative effects leaves the experiment before it is done. If we focus only on the remaining half, we would wrongly conclude that the treatment has overall positive impacts.

- attrition: Reduction in the size of samples over time, which, if not random, can lead to biased estimates.

Estimating causation with observational data

- data from randomized trials are not always available when important empirical questions need to be answered. Typically, what the analyst has instead are observational data.

- observational data: data generated by individual behavior observed in the real world, not in the context of deliberately designed experiments.

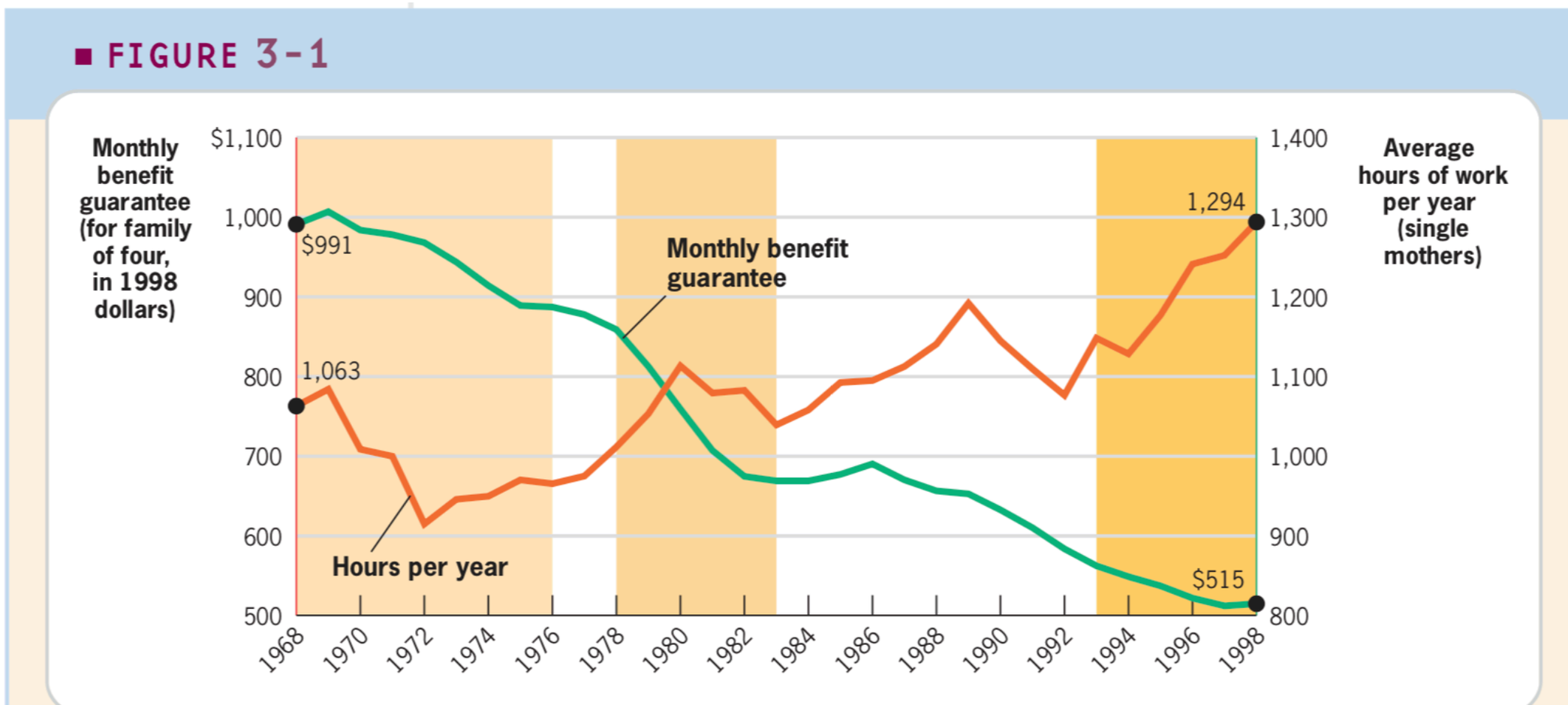
- For example, instead of information on a randomized trial of a new medicine, we may simply have data on who took the medicine and what their outcomes were.

- There are several well-developed methods that can be used by analysts to address the problem of bias with observational data, and these tools can often closely approximate the gold standard of randomized trials

- In other words, we can use observational data to estimate causal effects instead of just correlations.The major concern is how to overcome any potential bias so that we can measure the causal relationship (if there is one).

Time Series Analysis

- One common approach to measuring causal effects with observational data is **time series analysis**, the analysis of the comovement of two series over time.



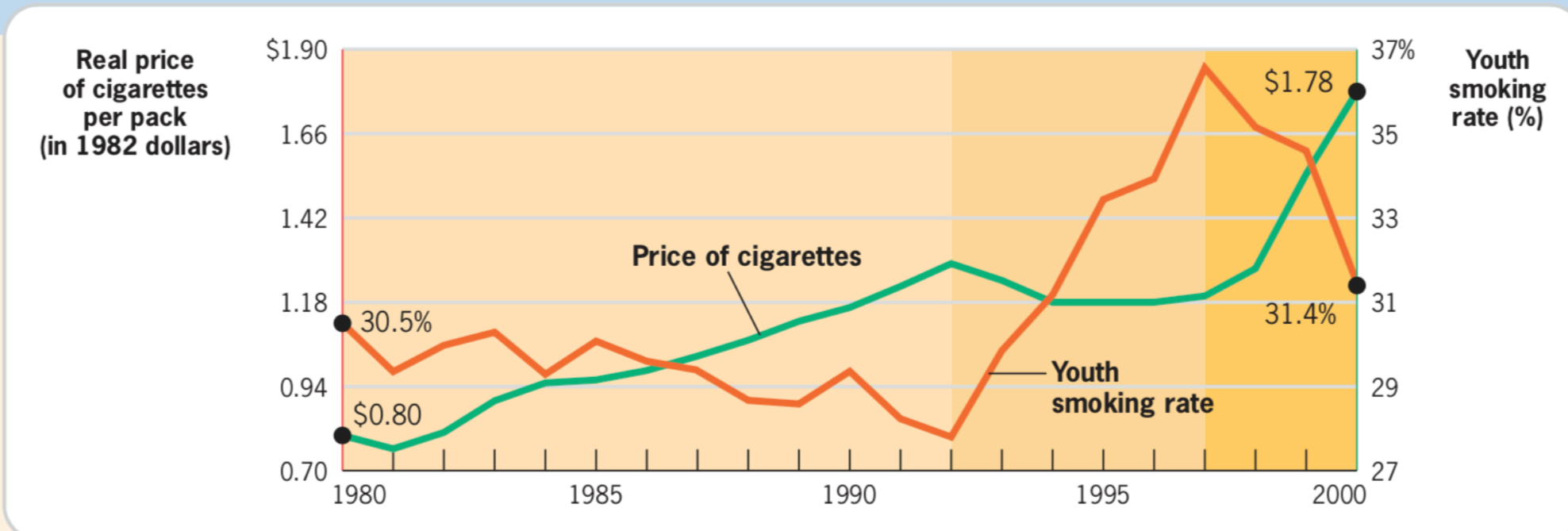
Average Benefit Guarantee and Single Mother Labor Supply, 1968–1998 • The left-hand vertical axis shows the monthly benefit guarantee under cash welfare, which falls from $991 in 1968 to $515 in 1998. The right-hand vertical axis shows average hours of work per year for single mothers, which rises from 1,063 in 1968 to 1,294 in 1998. Over this entire 30-year period, there is a strong negative correlation between the average benefit guarantee and the level of labor supply of single mothers, but there is not a very strong relationship within subperiods of this overall time span.

Time Series Analysis

- Problems with Time Series Analysis
  - Although the time series correlation can be striking, it does not necessarily demonstrate a causal effect. Other factors get in the way of a causal interpretation of the correlation over time; and these factors can cause bias in this time series analysis because they are also correlated with the outcome of interest.
- When Is Time Series Analysis Useful?
  - Is all time series analysis useless? Not necessarily. In some cases, there may be sharp breaks in the time series that are not related to third factors that can cause bias.
  - Thus, while time series correlations are not very useful when there are long-moving trends in the data, they are more useful when there are sharp breaks in trends over a relatively narrow period of time.

Time Series Analysis

■ FIGURE 3-2



Real Cigarette Prices and Youth Smoking, 1980–2000 • The left-hand vertical axis shows the real price of cigarettes per pack, which rises from $0.80 in 1980 to $1.78 in 2000. The right-hand vertical axis shows the youth smoking rate (the share of high school seniors who smoke at least once a month), which fell from 1980 to 1992, rose sharply to 1997, and then fell again in 2000 to roughly its 1980 level. There is a striking negative correspondence between price and youth smoking within subperiods of this era.

Source: Calculations based on data on smoking from Monitoring the Future survey and on tobacco prices from the Tobacco Institute.

Cross-Sectional Regression Analysis

- A second approach to identifying causal effects is **cross-sectional regression analysis**, a statistical method for assessing the relationship between two variables while holding other factors constant. By cross-sectional, we mean comparing many individuals at one point in time, rather than comparing outcomes over time as in a time series analysis.

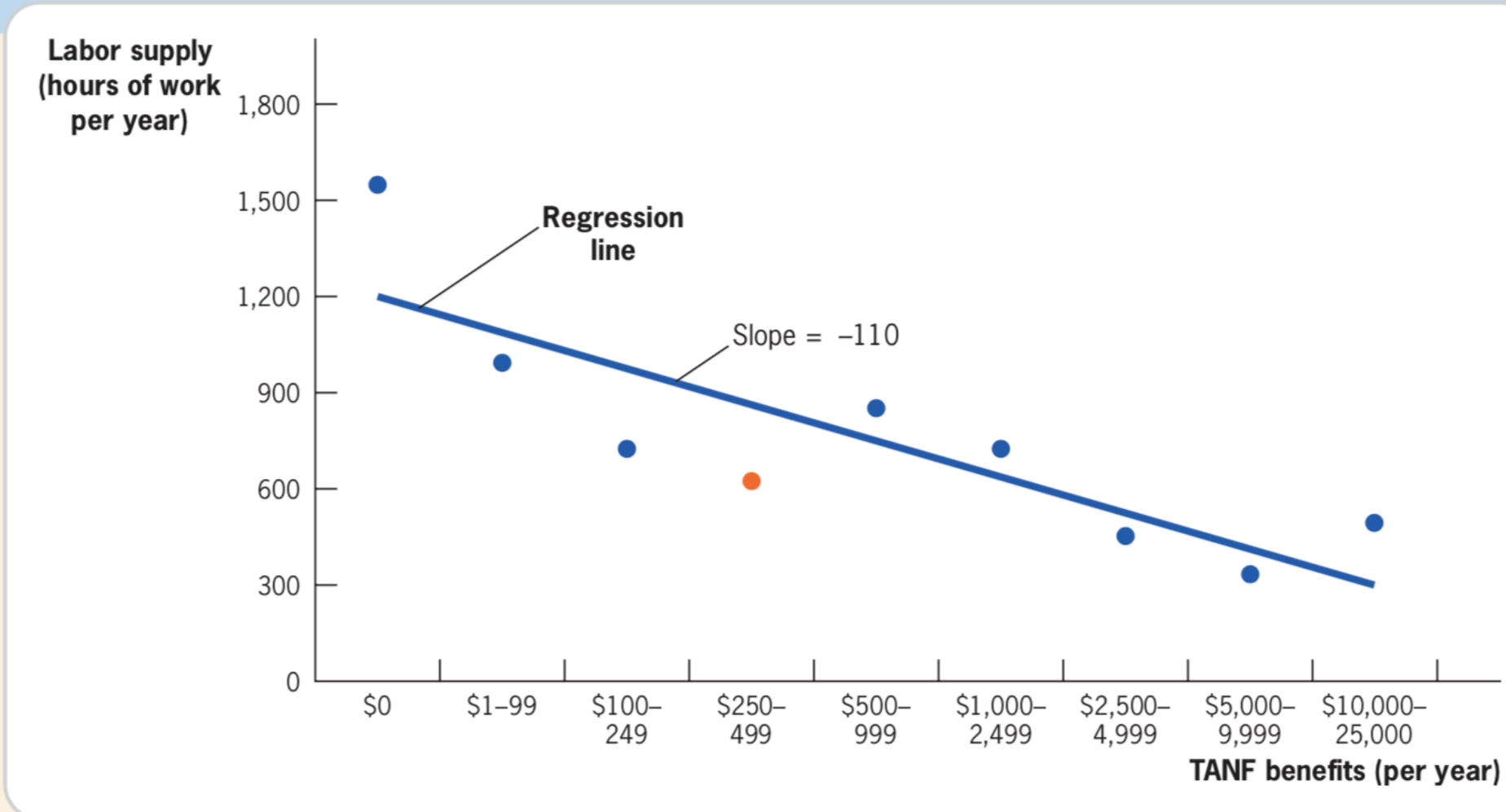
- Regression analysis describes (and quantifies) the relationship between the variable that you would like to explain (the dependent variable) and the set of variables that you think might do the explaining (the independent variables).

- The best approximation of such relationship is shown by the regression line. There is no single line that fits perfectly through this set of data points; instead, the linear regression finds the line that comes closest to fitting through the cluster of data points.

- Technically, this line is the one that minimizes the sum of squared distances of each point from the line. As a result, one major concern with linear regression analysis is outliers. An outlier, which is a point that is very far from the others, exerts a strong influence on this line, since we are minimizing the sum of squared distances, so a large distance has an exponentially large effect. For this reason, analysts often use other approaches that are less sensitive to such outlying observations.

Cross-Sectional Regression Analysis

■ FIGURE 3-4



TANF Benefit Income and Labor Supply of Single Mothers, Using CPS Data • Using data from the CPS, we group single mothers by the amount of TANF income they have. Those who are receiving the lowest level of TANF income are the ones providing the highest number of work hours.

Source: Calculations based on data from Current Population Survey's annual March supplements.

Cross-Sectional Regression Analysis

- The relationship between two variables approximated by the regression line is, again, not neccessarily causal. Therefore, we don't interrpret the results as "a x% reduction/rise in variable A is causing y% reduction/rise in variable B" but rather "a x% reduction/rise in variable A is associated with y% reduction/rise in variable B"

- Regression analysis has one potential advantage over correlation analysis in dealing with the problem of bias: the ability to include control variables. Control variables in regression analysis take into account other differences across individuals in a sample, so that any remaining correlation between the dependent variable and independent variable can be interpreted as a causal effect.

- However, in reality, control variables are unlikely to ever solve the problem of bias completely, as the key variables we want, are often impossible to measure in data sets. Usually, we have to approximate the variables we really want with what is available. These are imperfect proxies, however, so they don't fully allow us to control for differences.

Cross-Sectional Regression Analysis

- $y = \alpha + \beta x + e$
- where
  - α = constant (value for $x = 0$)
  - β = slope coefficient, represents the change y per unit change of x
  - e = error term, which represents the difference for each observation between its actual value and its predicted value based on the model

Table XXX: Linear regression analysis (Ordinary least squares regeression)

Dependent variable	(1) Total	(2) Charity	(3) Subjects	(4) U-supply	(5) O-supply
Part B – Part A (Simulations)	-1.57	-2.20	0.63	-0.02	-0.09
Wait	-56.14***	-67.13***	10.99***	0.45***	-1.65***
	(4.73)	(4.70)	(2.12)	(0.03)	(0.18)
Demand	4.12	0.77	3.36	-0.01	-0.40**
	(2.70)	(3.63)	(2.02)	(0.03)	(0.18)
Supply	7.11***	0.19	6.92***	-0.00	-0.75***
	(2.52)	(3.24)	(1.86)	(0.02)	(0.17)
Wait * Demand	-4.10	-0.78	-3.31	0.01	0.40
	(6.61)	(6.60)	(3.07)	(0.05)	(0.27)
Wait * Supply	7.73	13.57**	-5.84*	-0.09**	0.54**
	(5.60)	(5.61)	(2.96)	(0.04)	(0.26)
Subjects	-0.14	-0.58	0.44	0.00	0.06
	(3.36)	(5.06)	(2.70)	(0.03)	(0.23)
Wait * Subjects	-44.02***	-44.91***	0.89	0.30***	0.65*
	(7.54)	(8.24)	(3.66)	(0.06)	(0.35)
Demand * Subjects	-1.44	1.09	-2.53	-0.01	0.21
	(4.92)	(6.49)	(3.28)	(0.04)	(0.29)
Supply * Subjects	-12.62*	-9.80	-2.82	0.07	0.24
	(6.60)	(7.43)	(3.71)	(0.05)	(0.33)
Wait * Demand * Subjects	1.83	-3.85	5.68	0.03	-0.42
	(10.52)	(10.72)	(5.19)	(0.07)	(0.49)
Wait * Supply * Subjects	24.50**	21.38	3.12	-0.14	-0.26
	(11.84)	(13.53)	(5.74)	(0.09)	(0.51)
Constant	20.44***	21.78***	-1.33	-0.10***	0.03
	(3.44)	(3.92)	(1.60)	(0.03)	(0.15)
R^2	0.96	0.96	0.66	0.96	0.77
N	72	72	72	72	72
Seed dummies	Yes	Yes	Yes	Yes	Yes
p-values					
D = S	0.20	0.86	0.08	0.87	0.06
W*D = W*S	0.04	0.01	0.41	0.02	0.58
D*S = S*S	0.10	0.12	0.92	0.12	0.90
W*D*S = W*S*S	0.06	0.05	0.66	0.05	0.75

Note: Standard errors are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1%-level, respectively.

Quasi-Experiments

- As noted earlier, public finance researchers cannot set up randomized trials and run experiments for every important behavior that matters for public policy. We have examined alternatives to randomized trials such as time series and cross-sectional regression analysis, but have also seen that these research methods have many shortcomings which make it hard for them to eliminate the bias problem.

- Is there any way to accurately assess causal influences without using a randomized trial? Is there an alternative to the use of control variables for purging empirical models of bias?

- Over the past two decades, empirical research in public finance has become increasingly focused on one potential middle-ground solution: the quasi-experiment, a situation that arises naturally when changes in the economic environment (such as a policy change) create nearly identical treatment and control groups that can be used to study the effect of that policy change. In a quasi-experiment, outside forces (such as those instituting the policy change) do the randomization for us.

- With quasi-experimental studies, unlike true experiments, we can never be completely certain that we have purged all bias from the treatment–control comparison. Quasi-experimental studies use two approaches to try to make the argument that they have obtained a causal estimate.The first is intuitive: trying to argue that, given the treatment and control groups, it seems very likely that bias has been removed.The second is statistical: to continue to use alternative or additional control groups to confirm that the bias has been removed.

Quasi-Experiments

- Difference-in-difference estimator: The technique that tries to combine time series and cross-sectional analyses to address the problems with each. By comparing the change in population A to the change in population B, the estimator controls for other time series factors that bias the time series analysis within population A. Likewise, by comparing the change within each population, rather than just comparing the two populations at a point in time, the estimator controls for omitted factors that bias cross-sectional analysis across the two populations.

- Searching for a change in variable X
    - 2 periods (Y, Z)
    - 2 populations (A,B)
    - In period Y, the policy is the same for A and B
    - In period Z, there is new policy for A, while the policy for B is not changed
- x (population A, year Y) - x (population A, year Z) = Treatment effect + Bias
- x (population B, year Y) - x (population B, year Z) = Bias
- Difference = Treatment effect

Quasi-Experiments

■ TABLE 3-1

Using Quasi-Experimental Variation

Arkansas			
	1996	1998	Difference
Benefit guarantee	$5,000	$4,000	–$1,000
Hours of work per year	1,000	1,200	200
Louisiana			
	1996	1998	Difference
Benefit guarantee	$5,000	$5,000	$0
Hours of work per year	1,050	1,100	50

In Arkansas, there is a cut in the TANF guarantee between 1996 and 1998 and a corresponding rise in labor supply, so if everything is the same for single mothers in both years, this is a causal effect. If everything is not the same, we can perhaps use the experience of a neighboring state that did not decrease its benefits, Louisiana, to capture any bias to the estimates.

Structural Modeling

- The randomized trials and quasi-experimental approaches previously described have the distinct advantage that, if applied appropriately, they can address the difficult problem of distinguishing causality from correlation.Yet they also have two important limitations. First, they only provide an estimate of the causal impact of a particular treatment. Say that an experiment found that cutting variable A by 15% raised variable B rates by 4.5 percentage points. This is the best estimate of the impact of cutting variable A by 15%, but it may not tell us much about the impact of cutting variable A by 30%, or of raising variable A by 15%. That is, we can't necessarily extrapolate from a particular change in the environment to model all possible changes in the environment. These approaches give us a precise answer to a specific question, but don't necessarily provide a general conclusion about how different changes in something might affect behavior.

- The second limitation is that these approaches can tell us how outcomes change when there is an intervention, but often they cannot tell us why. We often care about the **structural estimates** of responses, the estimates that tell us about features of utility that drive individual decisions, such as substitution and income effects. Randomized or quasi-experimental estimates provide **reduced form estimates** only.

- Reduced form estimates show the impact of one particular change on overall responses. This second disadvantage of randomized or quasi-experiments is thus related to the first: if we understood the underlying structure of responses, it might be possible to say more about how people would respond to different types of policy interventions.

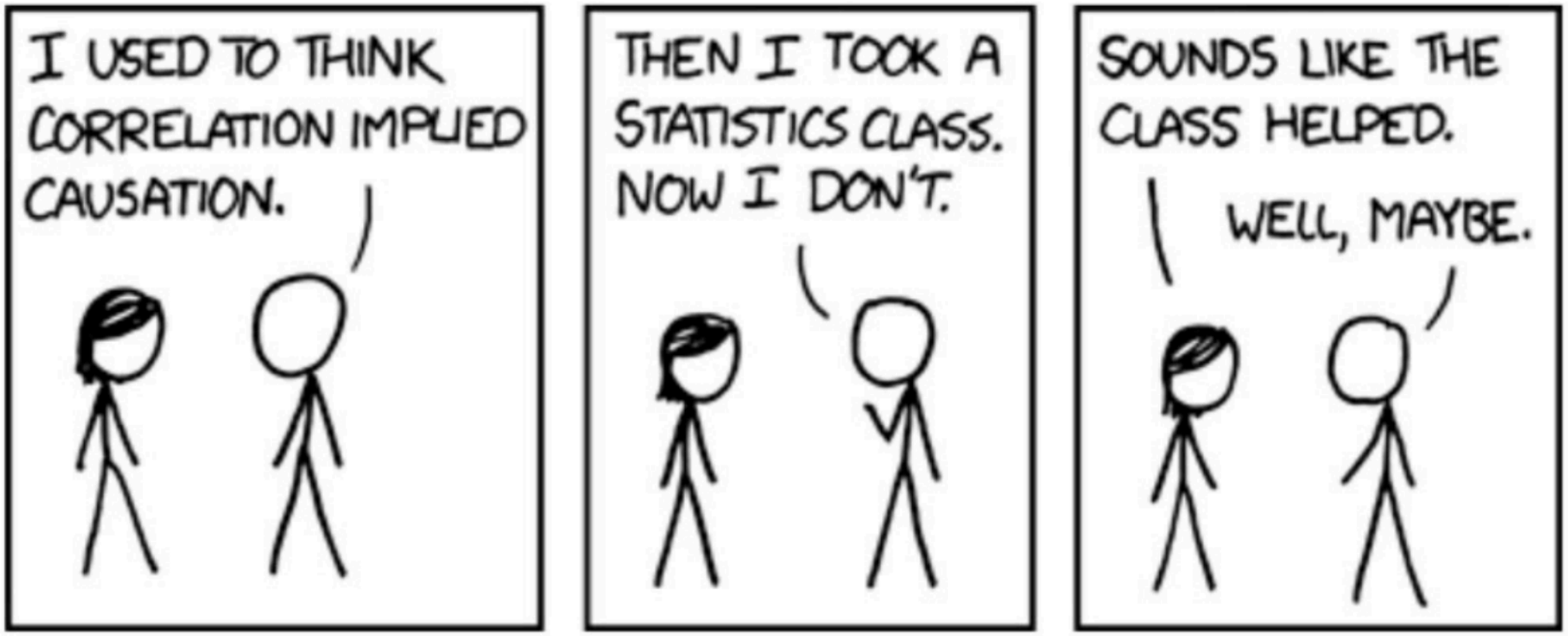
Structural Modeling

- These issues have led to the vibrant field of structural estimation. Using this research approach, empirical economists attempt to estimate not just reduced form responses to the environment but the actual underlying features of utility functions. They do so by more closely employing the theory to develop an empirical framework that not only estimates overall responses, but also decomposes these responses into, for example, substitution and income effects.

- Structural models potentially provide a very useful complement to experimental or quasi-experimental analyses. Yet structural models are often more difficult to estimate than reduced form models because both use the same amount of information, yet structural models are used to try to learn much more from that information.

Quantitative support for causation

- Model does not have to be truthful, but it has to be useful
- The more of the following conditions are met, the easier is to assign causation
    - Effect is significant - not only statistically significant, but also practically significant
    - Effect is consistent - same or similar results were achieved by other studies
    - Effect is specific - the factor of intereest influences only the response of interest, not also 10 other things
    - Effect follows time continuity - if x causes y, then x has to happen first
    - Effect is monotonous - more x causes more change in y
    - Effect is plausible - we know the mechanism behind the effect
    - Effect if supported by experimental results


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

More information

- Please have a look a this website, it is very nicely written and not too long or technical!

- https://towardsdatascience.com/everything-you-need-to-know-about-hypothesis-testing-part-i-4de9abebbc8a