Causal Inference and Understanding Causal Structure

Alex Wang

Professor Kevin Hoover, Faculty Advisor

Honor's Thesis submitted in partial fulfillment of the requirement of Graduation with Distinction in Economics in Trinity College of Duke University.

Duke University
Durham, North Carolina
2009

Acknowledgements

I would like to give special thanks to my advisor, Professor Kevin Hoover who took my shortcomings and successes with equal stride, and who struggled through the process alongside me.

But how will you look for something when you don't in the least know what it is? How on earth are you going to set up something you don't know as the object of your search? To put it another way, even if you come right up against it, how will you know that what you have found is the thing you didn't know?

- Plato's *Meno* (80D)

Abstract

This thesis aims to show that explicit understanding of possible causal structures often aids in inferring the true causes from data. This is done by first understanding that causes are chains of counterfactual dependence. Insofar as experiments, active or natural are not perfect, data can easily support false counterfactuals. Even those tools especially designed to identify unbiased estimates, like instrumental variables, often fail. Causal structure explains the failure of these tools, but more importantly allows us to better identify which counterfactuals to reject or accept.

I. Introduction

Causation is a vital area of study, particularly for the economist, because of the usefulness of correct causal inferences. Insofar as economics is a policy making tool, we must have correct causal inferences upon which we can build sensible policies. Pearl 2000 (pp. 331-358) provides a telling example. Although we have the more or less constant association of the rooster's crow signaling the dawn, it does not cause the sun the rise, thus it would be a foolish policy to make the rooster crow earlier as an attempt to shorten the night. Sound policy depends on identifying causes that are both real (as opposed to spurious) and relevant and appropriate (as opposed to those that we are not interested in). If we want to make the night shorter, identifying the rooster's crow would be useless because it does not actually causes the sun to rise, and identifying the earth's distance to the sun would be irrelevant because it causes the nights to be warmer, not shorter.

But how do we find the true, relevant causes? It helps, firstly, to know exactly what type of thing causation is. (Presumably, it is difficult to identify something without a good grasp of what it is.) Causes are derived from counterfactuals. More specifically, they are chains of counterfactual dependence. X causes Y (or in other words, there is a causal route from X to Y) just insofar as there is a chain of counterfactual dependence linking X to Y. Although this definition of causation is motivated by a long history of philosophical debate and a wide philosophical literature supporting it (see Menzies 2008), it takes only basic intuition to realize its veracity. Correct causal inference, is then just a matter of identifying the right counterfactual chain.

Various techniques for causal identification have arisen over the years, varying in complexity and elegance. They can more or less be identified with two types of approaches:

active experimentation and passive observation. Each approach has developed special tools to try to solve inferential problem of unobservables. But even these tools that by construction should be able to overcome the usual difficulties in inferring the causal route from *X* to *Y* from the data (namely active, randomized experimentation and instrumental variables) often still fail to identify the right chain of dependence. The unique nature of individual techniques can cause the identification of idiosyncratic causal routes. Often this can be preempted by recalling that causation is matter of chains of dependence and conjecturing about what structures of causal chains reasonably exist or might reasonably exist depending on the unique identification techniques applied.

No experiment is perfect, so inference from data alone will never guarantee unbiased results. It is the purpose of this thesis to show that what greatly aids in making causal inferences is an explicit understanding that causes are chains of dependence and how these chains affect the results we see. Such an understanding is useful in assessing the appropriate tools and what they can and will identify.

Given that the purpose of this thesis is to show the importance of understanding the true causal chain, I will make frequent use of causal diagrams, namely figures consisting of variables connected by arrows whose direction represent the direction of causation.

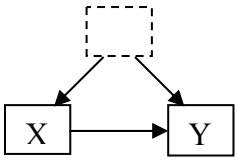

Figure 1

The diagram above states that *X* causes *Y* (but *Y* does not cause *X* because there is no arrow from *Y* to *X*) and that *C* is a common cause of both *X* and *Y*. Solid boxes will denote

observed variables and dashed boxes will denote unobserved variables. These diagrams will be used to represent the actual or presumed causal structures in actual or imagined situations to highlight how legitimate or spurious inferences are made.

The thesis will proceed as follows. Section II will give a general discussion of the difficulties in making correct causal inferences, namely that unobserved variables influencing the treatment variable and the outcome in question will bias inferences. Section III shows the conditions necessary to obtain unbiased inferences and how active randomization and instrumental variables satisfy these conditions. Section IV presents cases where the conditions described in Section III are clearly not enough to obtain unbiased inferences, citing two case studies from recent econometric literature. Section V first gives a general description of a specific context (self-selection) where randomization and instrumental variables may still fail to identify the sought after causal route, and then presents another aspect of one example from Section IV as a specific example. Section VI concludes by making the point that while we can add even more stringent conditions for unbiased inference, these newly discovered conditions, although sufficient, are not necessary in all circumstances to make valid inferences. These more stringent conditions are necessarily contingent on the causal structures that were realized. Moreover, we only discovered these conditions from knowledge of the underlying causal mechanism. Thus it is the causal chain that is primary. The right tools to use can only be understood in relation to the overall causal chain connecting the variables of interest.

II. Endogeneity

A. Causal Intuition

First we must understand what it means that causes are chains of counterfactual dependence. For our purposes, this does not require any great philosophical scrutiny, only enough to understand the basic principles that our everyday intuitions about causation entail, for these seemed have to served us well enough in everyday decision making. Additionally, the current program of causal analysis in philosophy is not to reduce causation to logical principles, but rather just to find a sound theory that matches up with our common intuitions.

Counterfactuals are statements of the form "If *X* had occurred then *Y* would have occurred" or "If *X* had not occurred then *Y* would not have occurred." [1] *Y* depends counterfactually on *X* if it stands in the relationships exemplified in the above counterfactuals statements. Lewis (1973) states that causal dependence is just counterfactual dependence when the counterfactuals refer to events (562-563). We can see from the rooster example that we know this intuitively. We say that the rooster's crow causes the sun to rise if and only if a change in the conditions of the rooster's crow produces a change in the conditions of the sun rising. But additionally, we say if the rooster's crow does cause the sun to rise, its absence would entail the absence of the sunrise. We are essentially saying two versions of the same counterfactual, namely "if the rooster had crowed, then the sun would have risen," and "if the rooster had not crowed, then the sun would not have risen." It seems that to say that something *X* causes something *Y* is just to that that *X* and *Y* stand in the counterfactual relationships exemplified by above. Lewis (1973) summarizes this point: "[…] we do know

[1] Counterfactuals, despite their name, "[…] need not be contrary [namely *counter*] to certain facts. [A counterfactual] is just a hypothetical" (Heckman, 2005, p. 2). The counterfactual statement "If *X* had occurred then *Y* would have occurred" is statement about the connection between *X* and *Y*, and this connection is true of false regardless of *X* does or does not happen to occur.

that causation has something or other to do with counterfactuals. We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it. Had it been absent, its effects – some of them, at least, and usually all – would have been absent as well" (p. 577).

This counterfactual understanding hints at the most basic conditions necessary to infer causation, namely that causal inference depends on a comparison of counterfactuals. Were we only to observe the first counterfactual of the rooster crowing and the subsequent association of a sunrise, we would not have license to infer that the rooster causes the sun to rise. We need to vary the rooster's crow to see whether the sun rises as a result. This helps us to reject certain spurious causations. If we say that the barometric reading on a certain barometer causes the air pressure, then adjusting the barometer should induce a change in air pressure. That no such change occurs suggest that the arrow of causation does not run from barometer to air pressure. This is just to say that changes in Y (variation in barometric readings) must counterfactually entail changes in X (variation in air pressure), or conversely that changes in X depend counterfactually on changes in Y.

While these examples show that counterfactual dependence is enough to imply causation, debates in philosophy show that a *single* counterfactual dependence is not enough. Lewis (1973) defines causation as a *chain* of counterfactual dependence, namely X causes Y if there is an event W between X and Y such that Y depends counterfactually on W and W depends counterfactually on X. Appeal of a chain of dependence is needed to solve problems where singular dependence implies faulty causal inferences, called problems of preemption in the philosophical literature (see Menzies 2008 for examples and how they are solved by appeal to the causal chain). I will not reproduce the problems here because they are not

problems generally encountered in economics. The move from defining causation as simple counterfactuals to chains of counterfactuals seems harmless enough because there should always be an event (maybe not an appreciably distinct event) between any two events. Additionally, distinct intermediate events should not be hard to find in economics because economics deals with events on the timeframes of human decision making. It is easy to imagine an event between completion of compulsory schooling at sixteen and earnings at thirty.

But for our purposes, the important part of defining causation as a chain of counterfactual dependence does not lie in whatever intermediary events exist between the variables we are interested in, but rather that it brings home the fact that causes are transitive. If X depends counterfactually on C and Y depends counterfactually on X, then Y depends counterfactually on C. If C causes X and X causes Y, then C causes Y. In the chain $C \rightarrow X \rightarrow Y$, X can be seen as our arbitrary intermediate variable W, so that the chain reduces to $C \rightarrow Y$. That every cause has an effect everywhere down the chain will be important to remember discussing the problems of causal inference, particularly selection bias.

B. Endogeneity

The simple inferential test of comparing counterfactual statements, though intuitively pleasing, meets two empirical problems. First, the data might not give enough variation in the treatment variable. If the rooster always crows at a certain time, we cannot infer from just this that it causes the sun to rise because we do not have the opposite counterfactual to contrast to. This problem is remedied insofar as we can find data with enough variation. It does not take much imagination to see that once we have observations of the rooster crowing

at different times, we will easily infer that it does not cause the sun to rise. But then we have the problem of endogeneity. This is the problem that since we do not observe *X* and *Y* in isolation, but rather in their natural circumstances, there may be other variables that may go unobserved (whether unaccounted for or even unobservable), that cause the systematic variation between *X* and *Y* from which we would infer a spurious causal link. These variables are said to be endogenous to the system *X* and *Y* because they influence the distribution of both variables. Figure 1 provides one example of how endogeneity can arise. *C* is endogenous because it is a common cause of *X* and Y.

Here is another example unlike the structure representing in Figure 1. Let *X* be the barometric reading and let *Y* be the onset of a storm. Both are caused by a change in air pressure, denoted C. Increasing air pressure increases the barometric reading *and* signals the onset of a storm. Decreasing air pressure has the opposite effects. The causal diagram is shown in Figure 2 below.

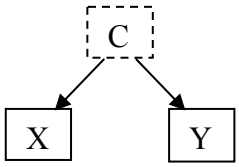

Figure 2

In this case *Y* depends counterfactually on *X*, namely if we see an increase in the barometric readings, then there is an increased likelihood of a storm and if we see a decrease in the barometric reading, then there is a decreased likelihood of a storm. In statistical terms, *X* and *Y* are negatively correlated, but as is well known, correlation does not imply causation. There is no arrow from *X* to *Y*. Yet we infer one from the counterfactuals supported by what we observe (namely the data). We observe only the values of *X* and *Y* which are the effects

of C, but because we do not observe C, we might assume that no other relevant causes are present, in which case the counterfactuals cause us infer a spurious causal connection.

Take an economic example. Let C represent innate ability which is unobservable and X represent the number of years of school one has attended and Y be the amount of one's annual earnings. If we wish to quantify the causal effect of schooling on earnings, for example the expected change in earnings due to a change in the amount of school attended, we will simply regress Y on X, assuming no other relevant causes. But the change in Y associated from a change in X that our regression provides us may be no more than just an association. X might not actually cause Y yet we might still observe from the data the counterfactual conditions that imply such a causal relationship. Variation in X is always associated with a systematic variation in Y and no variation in X, namely conditioning on X, will give no significant or systematic variation in Y. Given the data, we will reasonably say that "if X had not occurred then Y would not have" and "if X had occurred then Y would have as well," or rather, since X and Y are continuous variables "if X_2 had occurred instead of X_1, then Y_2 would have occurred instead of Y_1" where X_2 and X_1 are different levels of schooling and Y_2 and Y_1 are different levels of earnings.

This might happen because, as Angrist and Krueger (1991) propose according to Spence (1973), innate ability makes school less difficult and thus school attainment easier. Those with high innate ability obtain more schooling to signal to employers their higher ability, and those with higher ability also perform better and command higher earnings. Innate ability is what causes differences in earnings as well as differences in school attainment. But school attainment itself does not cause earnings. Schooling attainment serves as a proxy for ability, but were employers able to observe innate ability on its own,

then there would be no need for schooling, which served only as a signaling device. For example, those who have lower ability may force themselves through school, but employers being able to observe the level of their ability, will not offer them higher earnings. This is the counterfactual that "if one were to obtain more schooling (given that employers observe ability directly), one would *not* command higher earnings," namely that there is no causal connection between schooling and earnings.

We have assumed that there is no causal connection between X and Y, thus C being unobserved produces a spurious causal inference. But we can just as easily imagine that X does cause Y but that C is nonetheless unobserved and a common cause of them both (see Figure 1). In this case we would be correct in inferring that there that X causes Y, but we would be wrong in our inference of the strength of this relationship. In econometric terms, C being unobserved would bias our regression results. Schooling, independent of ability, might very well cause higher earnings because people obtain useful knowledge and skills that command higher salaries. But those with higher innate ability might stay school longer simply because they enjoy it and will obtain more skills than those with lower abilities. People that have obtained the same amount of schooling may still command different salaries because of differences in ability. The differences in earnings by level of schooling are now greater because of the influence of innate ability. Higher ability increases both schooling and earnings while lower ability decreases both. This is denoted in Figure 3 where the "+" signs indicate causes that promote the adjacent effect.

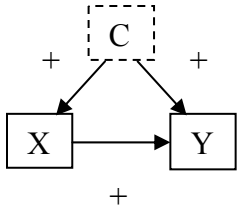

Figure 3

Without taking this into account, the causal effect of schooling on earnings from a simple regression of Y on X will be biased upward due to ability.

The general lesson to be learned from this discussion of endogeneity is that bias results from assuming that only the observable variables are the causally relevant ones when the data is generated from unobserved variables. *C* causes *X* and *Y*, namely that the variation in *X* and *Y depend* counterfactually on *C*, but given *C* is unobserved we cannot model this counterfactual dependence. We can only construct counterfactuals for variables we observe because counterfactuals require observable variation. (A counterfactual takes the form "if X_2 had occurred instead of X_1, then Y_2 would have occurred instead of Y_1.") If we assume that only *X* and *Y* are relevant for generating the data, then necessarily the data will support false counterfactuals linking *X* and *Y*. Namely, we generate the counterfactual "If X_2 instead of X_1, then Y_2 instead of Y_1" and assume it is true, that is we assume that the change in *Y* (Y_2 instead of Y_1) results from a certain change in *X* (X_2 instead of X_1). This counterfactual is false because it does not account for *C* which is unobserved. The true counterfactual is "If *C*, then if X_2 instead of X_1, then Y_2 instead of Y_1," namely the change is *Y* only results from such a change in *X* given the effects of *C*. This may be rewritten as "If X^*_2 instead of X^*_1, then Y_2 instead of Y_1" where X^*_2 and X^*_1 are values that have taken into account the endogenous effects of *C*. When we do not account for endogeneity, we are assuming that $X_2=X^*_2$ and $X_1=X^*_1$ when of course they do not. This is only true when there is no endogeneity, namely

only X and Y are the only causally relevant variables. A regression of Y only on X is the mathematical equivalent of this assumption. From it, of course the causal inference (the coefficient estimate from the regression) will be biased if not completely spurious. (Economists are of course not so naïve as to believe the assumptions of causal structure that the availability of data forces on them. They know that their causal inferences may be biased.) In order to obtain unbiased results, one must control for all endogenous variables, namely include the full causal structure. But such direct statistical control, namely including the variable in the regression, is not often feasible. Some variables, like innate ability, are not in principle measurable. It is often not possible to obtain data on all significant causal factors that might bias an estimate.

III. Active Experiments, Randomization, and Instrumental Variables

A. Active Experiments

Clearly endogeneity is a problem because it biases the results of our causal inferences. The bias is a consequence of the endogenous variable being unobserved and since regression equations can be constructed only on observed data, these will inevitably fail to reflect the true underlying counterfactuals. But this shortcoming, although presented in the context of passive observation in the examples above, is not a trait inherent to it. Active experiments do not solve the endogeneity problem. All active experimentation does is allow the experimenter to manipulate the treatment variable. The active experimenter can make the rooster crow by telling it a good rooster joke. She does not have to wait for the rooster to crow on its own at different times. But all this ensures is that counterfactuals involving the treatment can be constructed (the rooster crowed at 6 am versus at 5 am). Active

experimentation does not ensure that the *right* counterfactuals, namely those capturing the true causal mechanism, are constructed. Suppose, in studying the effects of schooling on earnings, one were able to compel people to obtain various amounts of schooling, namely one is able to actively vary the amount of treatment. Still, if the sample consists entirely of those with very high innate ability (because of selection bias), the results of such an experiment would still be biased. (The results of this experiment assumes that the counterfactual "if X_2 instead of X_1, then Y_2 instead of Y_1" is equivalent to the counterfactual "if ability is high, then if X_2 instead of X_1, then Y_2 instead of Y_1.") Like the passive analyst, the active experimenter must still control for an innumerable variety of possible confounding causes (those that might lead to selection bias), some of which are unobservable (like tastes or innate intelligence).

B. Randomized Experiments

To remedy this problem, scientists have employed randomized experiments, brought into prominence as a research method by Fisher (1926). The principle of the randomized experiment is that the variation in the treatment variable is produced by a random process. Being random, the treatment variable would necessarily be causally unaffected by any endogenous variable: "The key role played by the process of randomisation seems to be to ensure, up to a probability that can be calculated from the sampling distribution produced by the randomisation, that no uncontrolled common cause of both the treatment and the response variables could produce a spurious association. Fisher said as much himself when he stated that randomization 'relieves the experimenter from the anxiety of considering and estimating the magnitude of the innumerable causes by which his data may be disturbed" (Shipley, 200,

p. 9). This intuition is represented graphically in Figure 4 below, where *R* represents the random process by which treatment is administered.

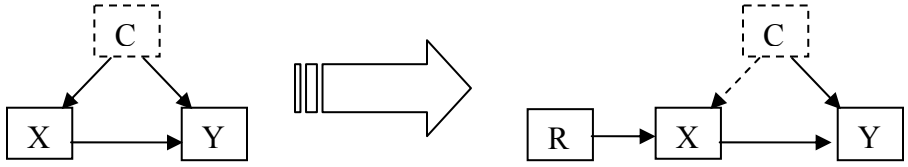

Figure 4

R is observable and because it is not caused by *C*, counterfactuals dependent on *R* represent the true causal structure. With randomization we know that the counterfactual "if R_2 instead of R_1, then X_2 instead of X_1" identified from the data is true, because there are no common cause of *R* and *X*. Using this, we also know that the counterfactual the counterfactual "if X_2 instead of X_1, then Y_2 instead of Y_1" is true because these variations of *X* are generated from *R*.

If *X* is the amount of schooling and is determined by a random process (say a random number generator with each number assigned to a specified amount of schooling) then innate ability will not bias the regression of *Y* on *X* because ability does not influence how much schooling a person receives, thus the arrow from *C* to *X* is dashed. (It is of course important to use only variation produced *solely* by the random process. Often all the variation is produced randomly, as with assigning all participants differing levels of schooling with a random number generator. When not all variation in treatment is caused randomly, then we must strip out and use only that portion is randomly determined to ensure unbiased results. This is done in econometrics by conditioning X on the random variable and regressing *Y* on this conditional.)

This of course is not to say that C no longer affects X in any way, just not in any systematic way, namely the *expected* change in X due to a change in C is zero. Each person who receives treatment, of course is still affected by C. If age, or height, or race were confounding, random selection do not make the participants ageless, height-less, or race-less, a tabula rasa on which only the treatment, when administered, exists. Rather these once confounding factors still exist, but just not so as to systematically affect the results of treatment. This is because we are constructing counterfactuals based on aggregate data, namely we are looking at average variations in the variables over the sample. Thus regression equations rightly express the *average* change in Y as given a change in X. The coefficient estimate that captures the causal effect is given by $E(Y|X)-E(Y|X)$. The average effect of C given randomization is zero. The situation would be different if we examined counterfactuals on individual people.

The point is that if innate ability were confounding because the population consisted of only those in higher levels of schooling that are not compulsory, namely only those with higher innate ability had self-selected into the sample we are looking at, randomly assigning treatment (levels of schooling) solves this endogeneity problem because random assignment, completely uncorrelated with innate ability, will presumably, evenly distribute treatment over a population of people of various innate ability levels, provided that the experimental population is large enough. Some will have higher abilities than others and perform better at work, thus commanding higher salaries, but these, solely ability-based performance levels will be uncorrelated with X and distributed evenly across the population, in essence balancing each other out, given a large enough sample size.

Randomized application of treatment is upheld as the acme of statistical achievement because the supposition of randomness is able to "break" the correlation of endogenous unobservables on the treatment variable. Moreover, randomization is able to break any endogeneity because correlation between variation in the endogenous variable and the randomly produced variation is necessarily zero. Not only is randomized treatment able to obviate the logical impossibility of conditioning on something that cannot be observed (like innate ability) but it also avoids the practical impossibility of conditioning on innumerable possible confounding causes.

Randomized treatment *identifies only* the causal arrow from treatment to outcome. With endogeneity, the variation in X and Y was in some unidentifiable part (because C is unobserved) determined by C. But with randomization, a random process generates variation in X, a variation which on average is *independent* of C. The variation in Y is still dependent on C, but these variations being uncorrelated with X will be differenced out – recall that the coefficient estimate is a difference of expectations of Y. Because of randomization, variation in X that is not "caused," namely systematically affected by C, can be identified and with this the true causal impact of X on Y.

C. Instrumental Variables

Although the randomized experiment is a statistical breakthrough that in principle greatly aids the program of causal inference, what we should notice is that randomization is *not* a necessary condition for resolving endogeneity. An endogenous variable is made no longer confounding so long as it is no longer correlated with the treatment variable. Randomized variation in the treatment variable is viewed as the best realization of this condition because randomness *credibly* ensures zero correlation. But the process determining variation in the

treatment variable need not be random, in so long as it is uncorrelated with endogenous unobservables and is itself observable.

Non-random processes may be used to infer unbiased estimates of the causal effect insomuch as they satisfy these conditions. Instrumental variables are examples of processes that satisfy these general conditions, namely they are variables that determine variation in the treatment variable and are uncorrelated with the endogenous unobservables. These two conditions are termed relevance and exogeneity respectively in econometric literature. Relevance is the condition that the instrument generates variation in the treatment variable. Formally, Z is a relevant instrument for X if $\text{cov}(X,Z) \neq 0$. Z is exogenous if it is uncorrelated with the error terms of Y, namely $E(Z|\varepsilon) = 0$, letting ε represent the error term. This entails that $\text{cov}(Z,\varepsilon) = 0$

The error term of Y is defined as the variation in Y that cannot be explained from the regression, namely that part of the variation in Y that is not caused by the variables one regresses on. For unbiased estimates of the effect of any X on Y, it must be assumed that $E(\varepsilon|X) = 0$. This is because a coefficient of X, which gives the effect of a change in X on Y is given as a difference of the expected values of Y given different values of X, namely $E(Y|X_2,C) - E(Y|X_1,C)$ where X_1 and X_2 are different values of X and C represents all other common cause of X and Y. If $E(\varepsilon|X) \neq 0$, namely the variation unexplained by the variables in the regression were not negligible, then the coefficient estimate that by definition cannot explain ε would be biased. The coefficient estimate cannot explain the ε-part of Y so it will be off by $E(\varepsilon|X)$, which is not zero. From our previous discussion, we know that biased estimates arises from endogenous unobservables, thus endogeneity is formally defined as $E(\varepsilon|X) \neq 0$. Notice that econometrically, the coefficient estimate is unbiased if $E(\varepsilon|X) = 0$,

namely the regression fully captures the true causal structure if there is no endogeneity – the counterfactuals constructible from the data represent the actual ones.

From the requirements of relevance and exogeneity, we know that a regression of Y on X conditional on Z, namely that part of X determined by Z will yield an unbiased estimate of the effect of X. This is because Z is exogenous to C, so $X|Z$ which is determined by Z is also exogenous to C. (The part of the variation in X that is determined by Z is not dependent on C because by definition, Z is causally independent of C – $E(\varepsilon|X) = 0$.) Regressing on this will then satisfy the non-endogeneity condition, namely $E[(X|Z)|\varepsilon] = 0$. This is analogous to what we have said about randomized experiments. The endogenous variables are uncorrelated with the random process that generates variation in X by its virtue of being random. Since treatment is determined entirely by this uncorrelated process, all variation in treatment is uncorrelated with any endogenous variables. With instrumental variables, the variation generating process (the instrument) is uncorrelated with the endogenous variables by definition. It does not determine all the variation in treatments, but it does determine some and that portion, $X|Z$, is uncorrelated with any endogenous variable, thus a regression only on this part gives an unbiased estimate.

To be clear, I have said that the instrument produces variation in the treatment variable, suggesting a causal relation. This is not guaranteed by the relevance requirement ($\text{cov}(X,Z) \neq 0$) because correlation does not imply causation. But correlation given the exogeneity condition will imply causation. This is due to the fact that while the converse is not true, causation typically implies correlation – except in the case of mutually inhibiting causes (Shipley, 2000, p. 9). Presumably, the economist chooses his instrument knowing something enough about its relation to X to know if it is such a case. If so, then know if there

is a causal relation and in all other cases, relevance is enough to entail that there is some causal mechanism connecting *X* and *Z*. If *Z* does not cause *X*, then *X* causes *Z* (either directly or indirectly). This possibility is given in Figure 5.

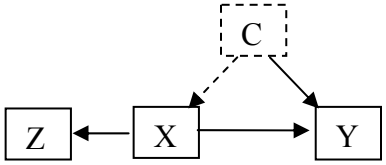

Figure 5

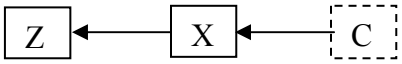

Figure 6

But since *C* cause X and *X* causes *Z*, then by transitivity, *C* indirectly causes *Z* (as in Figure 6), meaning that *C* and *Z* will be correlated violating the exogeneity requirement. Rather, if *Z* causes *X*, then *C* cannot cause *Z* so exogeneity is not violated. Another possibility, as shown in Figure 7, is given Reiss (2008, p. 135):

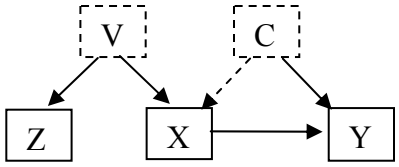

Figure 7

The instrument *Z* is not a cause of *X* but rather an effect stemming from an unobservable common cause *V* of *X* and *Z*. It should be noted that *V* is a valid instrument. It is uncorrelated with *C* and causes *X*. It is because it is unobservable that we must use *Z*, an unbiased proxy for *V*. (*Z* is an unbiased proxy because, like *V*, it is exogenous, so no there is no endogeneity to confound its relation with *V*.) Either way, an instrument entails an exogenous cause of the treatment variable. Most importantly, these are exactly the conditions that the random process in a randomized experiment satisfies. Thus instrumental

variables are just the passive counterpart to active randomization. Like randomized treatment, instrumental variables, being observable, identify only the causal chain from the instrument to *X* to *Y*.

IV. Identifying Too Much

Given that the subjects that economists study do not lend themselves to active control, the conditions that create instrumental variables are normally produced "naturally," namely without active manipulation. Thus the existence of instrumental variables signals a "natural experiment." Many "natural experiments" common in the econometric literature produce variation in the treatment variable from "[…] changes or spatial variation in rules governing behavior, which are assumed to satisfy the randomness criterion […] The major problem with these studies, however, is that the assumption of randomness is not credible" (Rosenzweig and Wolpin, 2001, p. 828). Because policy changes are of course man-made, there is a certain lack of credibility of their assumption of exogeneity. Policy makers may be seen as active experimenters, but even though they can actively vary the treatment, active manipulation alone does not guarantee exogeneity. Endogeneity may arise from various selection biases (Law makers make laws to please their constituencies.) There is another form of "natural experiments," namely natural "natural experiments" that obtain variation in the treatment variable from random variation in nature, in the forms of date-of-birth, twinning, and weather patterns. Because of this randomness, the instrumental variable assumption of exogeneity is more credible. But even so, we will see that random variation in the treatment alone may still not be enough to infer the correct causal effects. What is needed is a detailed understanding of the full causal chain from treatment to outcome, lest

our regressions model otherwise. To see this, we now turn to two applications of randomized "natural experiments" from the econometric literature, Angrist and Krueger (1991)'s quarter-of-birth experiment and Angrist (1990)'s Vietnam draft lottery experiment.

A. Quarter-of-Birth

Angrist and Krueger (1991) attempt to estimate the effect of schooling (actually compulsory schooling – but let us ignore this for now) on subsequent earnings. The authors employ three separate items that jointly induce variation in the amount of schooling attained, namely quarter-of-birth (what quarter of the year a person was born in), compulsory school entry laws, and compulsory minimum school leaving laws. The compulsory nature of the laws eliminates self-selection at entry, possibly due to unobservables such as ability, opportunity cost, or family circumstance. (Because we assume that the compulsory nature eliminates the selection effect of these variables, call them C, I do not bother to represent C in the causal graph). The fact that children are required to have turned six by January 1st of the year that they enter school but that students may leave as soon as they turn sixteen gives variation in schooling (Angrist and Krueger, 1991, p. 980). Those born in an earlier quarter-of-birth enter school at an older age but have the option to dropout at an earlier age than their counterparts in later quarter-of-birth cohorts. The innate ability levels, as well as any other unobservable confounding factors, are randomized across cohorts since quarter-of-birth is presumably uncorrelated with them. This setup is actually inherently susceptible to self-selection bias, which will be analyzed later. For now assume that the joint instrument, Z, produces no bias. We have seen how the instrument induces variation, thus it meets the requirement of relevance. Assuming no selection, it meets the requirements of exogeneity

because of the randomness of the quarter-of-birth component. The causal graph thus looks like the usual instrumental variable graph, shown in Figure 8, with μ representing the unobservable common cause of innate ability.

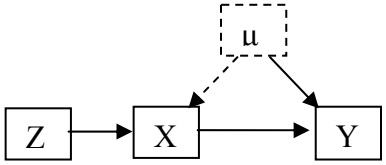

Figure 8

But as Rosenzweig and Wolpin (2000) point out, Angrist and Krueger (1991) do not include work experience in their regression. But work experience is perhaps just as, if not more important than schooling in determining earnings. Those with more work experience can more easily obtain higher paying employment than those with less work experience, thus work experience is another factor that determines earnings. The problem with the Angrist and Krueger (1991)'s omission of work experience is that conceivably, those that are not attending school are working. If there is a trade-off between attending school and working, then the years of school attended is negatively correlated with years of work experience, and this negative correlation is not relieved by the instrument. Variation in schooling is produced by the fact that those that are born earlier in the year enroll in school at a later age and thus are the first to reach the minimum school leaving age. Of the people that dropout upon reaching the leaving age, those born early (who leave first) have less schooling but, assuming that they work, will obtain more work experience. Thus quarter-of-birth induces variation in school attainment but also variation in negatively correlated work experience. Letting W represent work experience, the true causal graph looks like this:

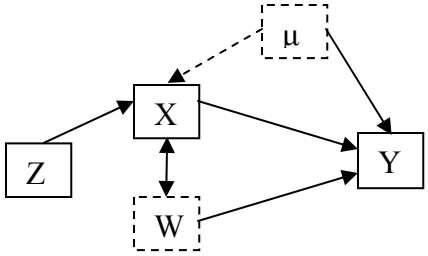

Figure 9

Because work experience is omitted, the estimate of the effect of *X* on *Y* is clearly biased, because more work experience means less schooling but also higher earnings. (I have chosen to use a double arrow between schooling and work experience not because of any concrete belief that schooling and work experience cause each other simultaneously, but just to emphasize the point that more schooling means less work experience and more work experience means less schooling. It is ambiguous which causes which. The point is that they are negatively correlated.) Those with less schooling will actually have higher earnings due to work experience than we would expect to see had we accounted for work experience.

What is important to note is that this endogeneity arises in spite of the randomized instrumental variable. Quarter-of-birth induces both variation in schooling as well as work experience, and insofar as work experience is omitted, quarter-of-birth is correlated with the error term through work experience, thus it fails the requirement of exogeneity. Here is an example of when randomness does not provide exogeneity. The instrument, being random, does obviate the endogeneity of ability because we have reason to believe why quarter-of-birth should be correlated with ability. But it is correlated with work experience through schooling. This shows that random instruments, though able to prevent some pre-existing bias in the system, do not prevent all of them. Angrist (1990) provides another example of instruments creating bias, but it is unique in that the instrument *creates* the bias. (Quarter-of-

birth does not create the bias because schooling and work experience would still be negatively correlated were no instrument used.)

B. Vietnam Draft Lottery

Angrist (1990) attempts to estimate the effect of military service on subsequent civilian earnings. The problem is that there are endogenous selection variables that affect voluntary military service as well as subsequent earnings. Angrist cites the possibility that men with fewer civilian opportunities are more likely to serve and that this lack of opportunity, perhaps due to lack of ability, would cause reduced earnings regardless of service. To control for such endogeneity, Angrist employs the Vietnam era draft lottery as a natural experiment. Random Sequence Numbers (RSN) from 1-365 were randomly assigned to birthdates in the cohort being drafted. Thus military service was determined by a random selection process.

Heckman in various papers objects that RSN is not a valid instrument for identification because of the unique procedure of the draft. Men only below a certain ceiling number were draft-eligible. Additionally, "Draft lottery RSNs were randomly assigned in a televised drawing a few months before men reaching draft age were to be called. Draft eligible ceilings […] were announced later in the year, once Defense Department manpower needs were known" (Angrist, 1990, p. 314). Heckman (1997) notes that "The higher the number in the draft, the less likely was a person to be drafted" (449). This is due to the presence of a ceiling. This coupled with the fact that there was a lapse in between the realization of the likelihood of being drafted with actually being drafted gave men of each cohort (and possible employers) time to make human capital investment decisions based on the likelihood of being drafted. This process is given by the Figure 10.

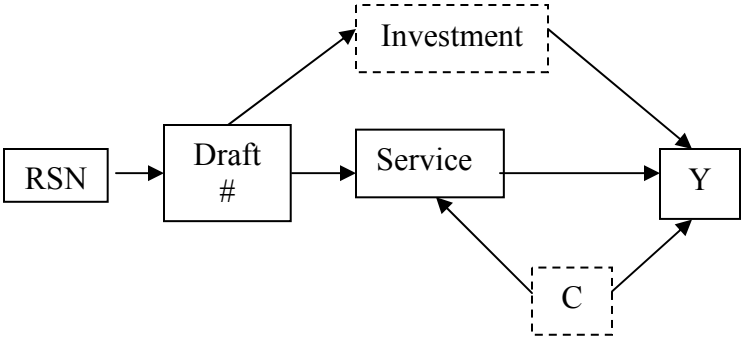

Figure 10

Because induction into service through the draft lottery was random, mandatory, and enforceable, the causal arrow from unobserved endogenous variables, *C*, is literally broken. The problem is that the draft itself produces endogeneity in the form of human capital investment. Earnings surely depend on the amount of human capital obtained. Were there no delay between assignment of draft numbers and the announcement of the draft ceiling, namely that people were immediately drafted, RSN would presumably identify the effect of service on subsequent income. Given the actual draft process, RSN identifies not just the effect of service on earnings, but service and investment due to the unique draft process. This added effect is different from very probable subsequent endogeneity after service because such bias would still be a result of service. We might reasonably assume that subsequent endogenous unobservables would result in most cases of military service, in which case the coefficient on RSN would not be problematic from a policy making standpoint with regard to service because the results would be robust for different instances of military service.

The problem as it stands is that the endogeneity is caused by the unique draft process and thus would not be robust for other forms of inducement into military service. Again, we have an example of completely randomly induced variation still being unable to identify the

right causal chain. The causal chain produced by RSN terminates in earnings via more than just military service because of self-selection.

V. Selection

A. General Framework for Selection

Although the selection that biases the estimate in Angrist (1990) is a function of the unique nature of the instrument, we know from experience that self-selection is a general problem that is inherently hard to cure. Rational people will self-select based on any available information to their benefit, and there will presumably always be an opportunity for self-selection insofar as treatment is not compulsory, which maybe most of the time. Disregard for human consent is usually unethical. If there is no way to preempt selection, the only other alternative to recover the estimate we want, namely the estimate on the whole population and not the self-selected one. We might think to do this by employing another instrumental variable, perhaps one that employs a random process to re-determine eligibility for treatment. This approach may be based on the fact that instruments identify a causal route starting from the instrument and going from X to Y and that since the instrument is exogenous, the causal route will not be biased. We shall see that while an unbiased causal route is identified, it is not the same one were there no self-selection. Valid instruments, while being able to prevent some endogeneity, certainly cannot undo them.

To see this, consider what may be termed the "mean treatment effect" or the "average treatment effect," which Heckman (1997) describes as "the effect of 'randomly assigning a person in the population to the program.' That counterfactual is $E(\Delta|C) = E(Y_1-Y_0|C)$" (443), where Y_1 is the outcome given participation in the program (receiving the treatment X), Y_0 is

the outcome without participation, and *C* is a set of covariates that determine *Y*. $\Delta = (Y_1 - Y_0)$ is the gain from participation in the program. This estimate is intuitively appealing because it gives the effect of treatment (denoted by the subscript on *Y*) and is indeed the coefficient of a regression of treatment (say *X*) on *Y* assuming no endogeneity other than *C*, which is conditioned for.

However, one source of endogeneity that may go unaccounted for is self-selection bias. Self-selection can occur at the point when treatment is accepted but after when treatment is first offered. We may wish to condition on *C* because it affects the treated population when the treatment is first offered as well as Y. In a clinical trail for a new pharmaceutical drug, *C* may be the demographic backgrounds that one typically controls for (through randomization). However, even given these controls, we may expect that people are well informed of their idiosyncratic, side-effects ahead of time (from previous experience) and that only those that experience mild side-effects decide to take the treatment. If *Y* is the biological effects of drug, then clearly *X* is correlated with *Y* (letting side-effects be the error term for *Y*) through the selection process (*S*).

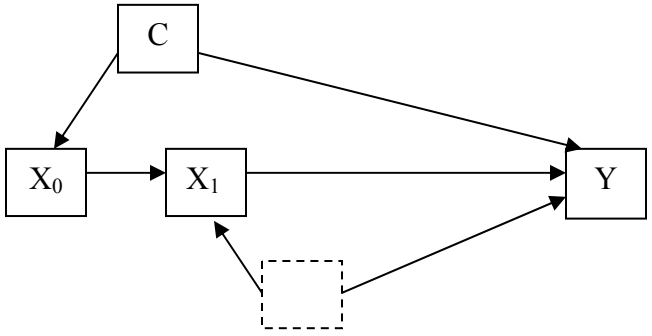

Figure 11

Let X_0 take on values for treatment at the time when treatment is first offered and X_1 represents the treatment at the time those who have volunteered for treatment and are ready

to receive it. In order to break the endogeneity caused by S, one might propose a random process (like flipping a coin), denoted by *R*, to see who receives treatment and who does not, namely to determine those who are eligible. Assume all who are eligible will actually receive treatment. (This assumption seems unproblematic since the population that is randomized consists only of those that want to receive treatment because they we suffer lighter side-effects.) *R* causes X_2, the receipt of treatment *after* selection.

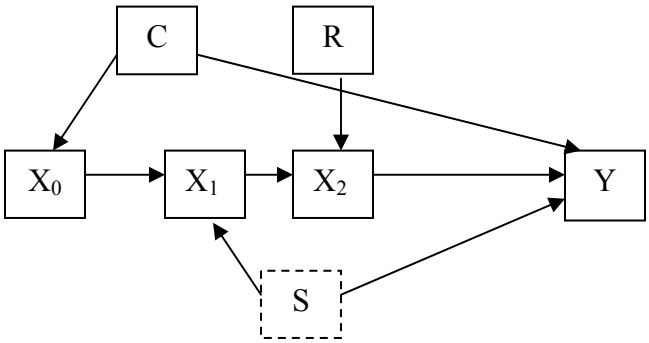

Figure 12

But irrespective of randomization, treatment has already been assigned to a self-selected population that is prone to lighter side-effects, namely one in which mean outcomes will be different from the mean outcome of the original population. The coefficient identified by instrumenting on *R*, namely the coefficient on $(X_2|R)$, is what Heckman terms the "mean effect of the treatment on the treated." It is given by $E(\Delta|D=1,C) = E(Y_1-Y_0|D=1,C)$" where $D=1$ for persons that are treated. ($D=0$ for those that are untreated.) I do not reprint the derivation here, but (Heckman 1997) shows that this effect differs from $E(\Delta|C)$ by $E(U_1 - U_0|D=1,C)$, where U_1 and U_0 are the error terms of *Y* for the treated group and untreated groups respectively.

We might expect that since $S \rightarrow X_1 \rightarrow X_2$ that an exogenous variable R would eliminate the influence from S. One might assume this because the causal route from selection to the instrument in Figure 12 can be rewritten as shown in Figure 13 because S causes X_2 through X_1.

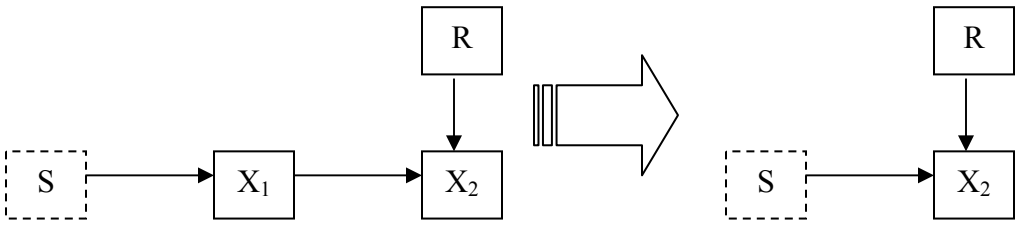

Figure 13

This then looks like the same setup as we normally see when instrumenting to try to overcome endogeneity. Thus we might assume that the effect of X_2 on Y should not be influenced by S and that we have instrumented the selection out.

It is true that we instrument out the effects of S on X_2, namely a regression of Y on $X_2|R$ will be unbiased. But this is not to say that a regression of Y on D (treatment or not someone gets treatment) will be unbiased. X_2, we will remember, is treatment after selection, namely treatment *given* that one wants treatment. If we do not instrument X_2 with R, then we know that $\text{cov}(X_2, U_1) = 1$, namely those who want treatment get treatment. Also $\text{cov}(X_2, U_0) = -1$, namely those that want treatment will not get treatment. But if R determines who gets treatment, than there will not be any systematic relation between X_2 and the errors. For those who want treatment it is random whether they will receive it or not. Thus R resolves the endogeneity of S on X_2.

But it does not resolve the endogeneity of S on D, namely there is still correlation between getting treatment and wanting treatment, namely in this setup, one must want

treatment before having the possibility of getting it. Obtaining treatment *depends* on wanting it. (D can be seen as the unconditional form of X_2. X_2 represents getting treatment *given* one wants treatment, while D only represents getting treatment.) This is because although R is determines the distribution of X_2, it does not solely determine the distribution of D (whether someone receives treatment). By the point we get to X_2, D is determined by a two stage process. Let $S = 1$ if a person allows herself to be treated, and 0 otherwise. Let $R = 1$ if for a person if they are offered treatment and 0 otherwise. Assuming that all people who seek treatment always accept if offered, at X_2, $D = 1$ only if both $S = 1$ and $R = 1$. $D = 0$ if either $S = 0$ or $R = 0$. Thus at X_2, Pr(D=1) = f(S,R) while Pr(D=0) = g_1(R) or g_2(S). To see explicitly why D conditional on R cannot recover the "average treatment effect" (ATE) but can recover the effect of "treatment on the treated" (TOT), let us make use of Heckman (1997)'s derivation. Heckman (1997) shows that a regression of D (whether or not an individual receives treatment X), given no assumptions about whether or not receiving treatment D is correlated with the errors, can be written:

$E(Y_1 - Y_0|C) + E(U_1 - U_0|C, D=1) + [E(U_0|C, D=1) - E(U_0|C, D=0)]$

Writing each expectation as a weighted sum, with conditioning on C implicit, we have

(1) $E(U_0|D=1) = (U_0) \cdot [Pr(U_0,D=1)/Pr(D=1)] = (U_0) \cdot [Pr(U_0,S=1,R=1)/Pr(D=1)]$

$= (U_0) \cdot 0 = 0$

(2) $E(U_0|D=0) = (U_0) \cdot [Pr(U_0,R=0)/Pr(D=0)] = (U_0) \cdot [Pr(U_0)Pr(D=0)/Pr(D=0)]$

$=(U_0) \cdot Pr(U_0) = E(U_0|C) = 0$

(3) $E(U_1 - U_0|C, D=1) = E(U_1|D=1) - E(U_0|D=1) = U_1 \cdot [Pr(U_1,D=1)/Pr(D=1)]$

$- U_0 \cdot [Pr(U_0,D=1)/Pr(D=1)] = U_1 \cdot [Pr(U_1,S=1,R=1)/Pr(D=1)] - 0$

(1) follows from the fact that it is impossible for a treated person to obtain an outcome associated with the untreated group. (2) follow from the fact that the value of D=0 is determined *solely* by a random event, R, that is independent of the value of the errors (which are functions of S because they represent the effect of unobservables) and that the expected value of the errors given C is zero by definition. (1) and (2) form the mean selection bias, which is zero when instrumenting on R. (3) is the difference between the TOT and the ATE, which is not necessarily zero from instrumenting on R. This is due to the fact that D=1 is still a function of S=1, which influences U_1.

Recall that to identify unbiased estimate of the effect of D on Y from the data is for the counterfactual "if D=1 instead of D=0, then Y_1 instead of Y_0" to be true. We know that it might not be true because of S, so we appeal to an instrumental variable, R. R identifies the true causal effect of D on Y if R can produce the values of D exogenously, namely if the counterfactual "if R=1 instead of R=0, then D=1 instead of D=0" is true. But this counterfactual is not true because D=1 depends counterfactually (namely is caused by) S and R. While one half of the dual counterfactual is true, namely "if R=0, then D=0" is true, the other half is not, namely "if R=1, then D=1." Rather the true half is "if R=1 *and* S=1, then D=1." R alone cannot identify the full counterfactual that quantifies the causal effect of D on Y. We need certain, exogenous variation in D to see an unbiased causal effect. R provides enough variation for us to observe an effect, the TOT, but since part of it is not exogenous, instrumenting on R does not give an unbiased effect, the ATE.

The general lesson is that there is a difference from identifying an unbiased causal route and the relevant or sought after, unbiased causal route. We know the instrument R identifies an unbiased causal route because X_2 is determined by a random process and thus is

uncorrelated with the error term. But this causal route gives us the effect of treatment on those that wish to be treated. This is not relevant if we wish to identify the average effect of treatment, namely the effect on any randomly chosen person. This is akin to the example given in the introduction. If we wish to make the night shorter, we want to know what causes the duration of the night, not what causes its temperature. Sometimes we need results unbiased by selection.

B. Angrist and Krueger (1991) revisited

The previous discussion assumes selection before treatment. Randomized treatment after selection has occurred cannot identify unbiased results. One way to eliminate self-selection bias, from an observational standpoint, is to rule out that any occurred, namely make sure that participants did not infer any idiosyncratic benefit from treatment. Heckman correctly cautions that such knowledge is most likely unavailable to the econometrician. From the standpoint of active experimentation, one could refuse to tell participants what the treatment was, so that they could not be able to infer any benefit. However, under this scenario, it would be doubtful that anyone would take the risk to participate, and those who did would automatically reveal their lack of risk aversion that maybe correlated with the outcome.

It seems that the only way to eliminate self-selection bias is to preempt self-selection. To estimate the "mean treatment effect," we have to instantiate the counterfactual it represents, namely randomly assign people to treatment and make sure that they take it. Disregard for consent in human test trails is usually unethical, but there are some exogenous events that may prevent self-selection into treatment such as the institution of compulsory laws. There is always the concern that lawmakers make laws aimed to please their constituencies, so that cross state/constituency variation will not be informative. The other

worry is that within state data supplies too little variation to make statistically credible inferences. National laws, then would not suffer from this problem.

We now return to Angrist and Krueger (1991), ignoring their omission of work experience, because their instrument employs the compulsory school entry law, which being compulsory, would prevent self-selection out of school. We see, however, that insomuch as there is selection into treatment, there is selection out of treatment.

To understand the instrument's failure despite its compulsory nature, let us look at the framework developed by Rosenzweig and Wolpin (2000). There are two types of people, each with varying ability. Type 1 will finish school while type 2 will drop out as soon as they reach the mandatory minimum school leaving age (a_k). Individuals decide whether to attend school only for one period beyond the school leaving age. "School attendance in the period following a_k, period one, is denoted by s_1=1 and nonattendance by s_1=0; completed schooling, at the end of period one, S_1, is therefore either S_0+1 or S_0" (834). Ignoring work experience (as Angrist and Krueger do), let μ=ability, and π_1 the proportion of type 1 in the population, and π_2 (=1- π_1)the proportion of types 2's. We are looking at the effect of *S* on the log of wages (ln(y)). The expected value of log wages is given by

$$E[\ln(y)] = \pi_1[f(S_0+1, \mu_1)] + (1- \pi_1)[f(S_0, \mu_2)]$$

Now "consider reducing the age at entry by one year, leaving the minimum school leaving age unchanged. In that case, both ability types will complete S_0+1 years of schooling; type 1's do so because it is optimal, while type 2's are forced to remain in school an extra year. The difference in expected (ln) incomes associated with the alternative school entry ages divided by the corresponding difference in expected schooling levels, that is, the Wald estimator, is $f(S_0+1, \mu_1) - f(S_0, \mu_2)$, which is the marginal return to schooling of the less able

type [type 2]" (Rosenzweig and Wolpin , 200, p. 836). The Wald estimator of the effect of *S* on *Y* is given by

$\Delta E[\ln(y)]/\Delta S$

The change in the expected value of log wages due to a change in age at entry by one year is

$$\Delta E[\ln(y)] = \{\pi_1[f(S_0+1, \mu_1)] + (1- \pi_1)[f(S_0+1, \mu_2)]\} - \{\pi_1[f(S_0+1, \mu_1)] + (1- \pi_1)[f(S_0, \mu_2)]$$

$$= (1- \pi_1)[f(S_0+1, \mu_2) - f(S_0, \mu_2)]$$

The change in *S* reduces to

$$\Delta S = [\pi_1(S_0+1) + (1- \pi_1)(S_0)] - [\pi_1 f(S_0+1) + (1- \pi_1)[f(S_0+1)]$$

$$= 1-\pi_1$$

Thus the Wald estimator is

$$\Delta E[\ln(y)]/\Delta S = f(S_0+1, \mu_2) - f(S_0, \mu_2)$$

Rosenzweig and Wolpin (2000) point out that this is "the marginal return to schooling of the less able type" (p. 836). This is true as each part of the equation is a function of μ_2, the ability of type 2's only. Similar manipulations will show that keeping the mandatory entry age constant and increasing the minimum leaving age, provided that it is less than S_0+1, will also yield an estimator that is a function only of type 2's. This is due to the fact that the true causal graph looks like this

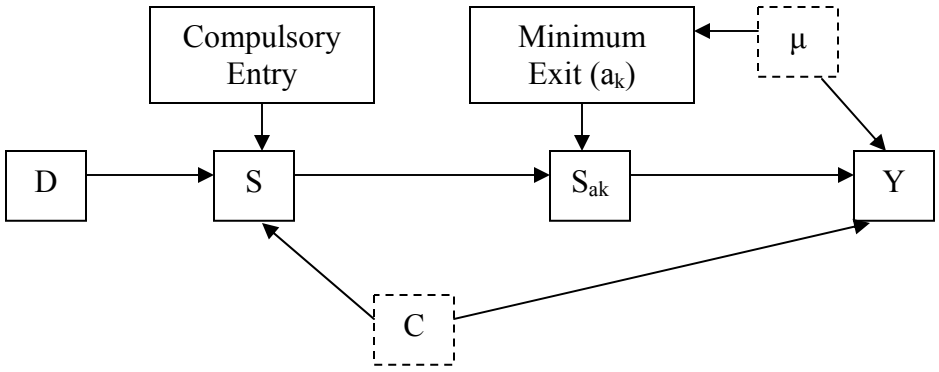

Figure 14

Although compulsory entry, because it is compulsory, breaks the endogeneity caused by C, unobservable innate ability is still problematic because it is correlated with schooling through the minimum exit age and to Y. This is because those who are affected by the minimum leaving age are those who wish to leave, and this is presumably highly correlated with ability. Quarter-of-birth and compulsory entry age may not be correlated with ability, but minimum leaving age is and all three are needed to produce variation to estimate an effect. Notice that varying either entry age or exit age will lead to a biased estimate because both are needed estimating the effect. Also notice that while our general example of selection led to only recovering TOT because of selection into treatment, here we have selection out of treatment and thus another example of recovering TOT, except that it is the effect of treatment on those that would prefer to be untreated. What is needed here is of course exogenous variation in access to schooling that affects the entire population, such as laws eliminating education across the board (perhaps from an extreme and sudden state insolvency) or a natural "natural instrument" such as a natural disaster that disrupts schooling. Here the problem will be that the instrument will not be exogenous to something like earnings.

Here the argument may be made, as it has been variously by Heckman that TOT is unproblematic because it is a more natural estimator than ATE. For example compulsory laws are targeted towards those that would act otherwise, so it is natural that varying these laws gives us only their effect on the group that they were meant to treat. What is unnatural is to force a treatment on someone who does not need it through random selection from the entire population. The problem with this objection is that it assumes certain determinism in human decisions. People make decisions based on circumstance and information, both of which are subject to change. Granted in the quarter-of-birth example, endogeneity is caused by innate ability that is presumably not subject to change, but other circumstances such as socio-economic situation are. Insofar as we want robust estimates over socio-economic status, we will have to be able to measure the ATE. Additionally people may base their decisions on asymmetric information. Only looking at the effect of a group selected on incomplete information will necessarily limit the results to that specific causal process. The Vietnam draft lottery is an extreme example.

Of course with observational data, we can only look at realized events, but some realized causal chains (namely natural experiments) provide suitable counterfactuals to other causal chains. Since any variable can cause selection bias, the realized treated populations may be very unique, thus efforts should be taken to extrapolate from them. The way the world usually is, is based on current circumstances, but these circumstances may change, and for the better (people may get access to more information or socio-economic situations may balance out). These changes, however, do not always invalidate the question of the effect of treatment. The effect of schooling on future earnings will still be relevant if all people had

the same ability or the same socio-economic status. In this way, we see that although selection gives us specialized situations, the general effect should not be ignored.

This aside, we see another example of randomness, although itself exogenous, ultimately failing identify an exogenous causal route. Certain steps can be taken to try to ensure that such a route is identified. Since there can be selection in as well as selection out, one may try to make any variation in treatment the result of compulsion. The chain of dependence to the outcome of interest may be long (in the case of adult earnings) so we must be wary of possible detours off the stated path. As a general rule, whenever possible, shorten up the chain of dependence to prevent detours and selection, as seen from the Vietnam draft example.

VI. Conclusion

The most forceful impact of these examples, however, is not given by the specific caveats that they demonstrate. Rather, it is that even these natural "natural experiments," representing some of the most fortunate and cleverly conceived situations that any passive analyst would hope to find, are still very prone to lead to false inferences. It seems that data, no matter what conditions generated it, can almost always be used to instantiate false counterfactuals. Something else must be appealed to, namely explicit *a priori* causal beliefs. At times, even light speculation about the causal structure will obviate some inferential problems. Clearly work experience is related to schooling and earnings. Other times, more in depth considerations must be made into the unique nature of the specific program of identification. In either case, causal beliefs about the chain of counterfactual dependence must be appealed to.

In the world of actual experience, no real experiment, natural or man-made is perfect, so there will always be opportunity for faulty inference from the data alone. Consider a famous thought experiment from Friedman and Schwartz (1963).

To answer the question of the stock of money causing business rather than business causing stock of money, authors employ a comparison to an alternate theory, namely a pin theory of business cycles. Suppose, they say, that "one could marshal a similar body of evidence demonstrating that the production of dressmakers' pins has displayed over the past nine decades a regular cyclical pattern" (48), such all the evidence they had just presented to support a money theory of business cycles could just as easily and faithfully be reproduced with pins replacing money. Still no one would accept a pin theory of business cycles. Why not? Because:

> We know that while pins are widely used and occasionally of critical importance, taken as a whole, they are a minor, if not trifling, item in the economy. […] We can readily conceive of an economy operating without pins yet experiencing cycles like those of history; we can readily conceive of large autonomous changes occurring in the production of pins, but we cannot readily conceive of any channels through which such autonomous changes could have wide-reaching effects on the rest of the economy (p. 49).

Simply put, our economic theory, well understood and thoroughly convincing, would not allow a causal structure that would give rise to a pin theory of business. Friedman and Schwartz (1963) spend the rest of their paper explicating the probable causal structure based on theory. Beliefs about the causal chain of dependence, based on theory, are primary. These beliefs are helpful because they allow us to think speculate which counterfactuals to instantiate and test, but more importantly, which counterfactuals to accept or reject.

References

Angrist, Joshua. (1990). "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence From Social Security Administration Records." *American Economic Review* 80 (3): 315-335.

Angrist, Joshua, and Alan B. Krueger. (1991). "Does Compulsory Schooling Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106 (4): 979-1014.

Angrist, Joshua, Guido W. Imbens, and Donald B. Rubin. (1996). "Identification of Causal Effects Using Instrumental Variables." *Journal of American Statistical Association*. 91 (434): 444-455.

Bronars, Stephen G., Jeff Grogger. (1994). "The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment." *The American Economic Review* 84 (5): 1141-1156.

Heckman, James. (1997). "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 32 (3): 441-462.

———. (1996a). "Randomization as an Instrumental Variable." *The Review of Economics and Statistics* 78 (2): 336-341.

———. (1996b). "Identification of Causal Effects Using Instrumental Variables: Comment." *Journal of American Statistical Association*. 91 (434): 459-462.

———. (2005). "The Scientific Model of Causation." *Sociological Methodology* 35: 1-97.

Hoover, Kevin. (2001). *Causality in Macroeconomics*. Cambridge: Cambridge University Press.

——. (2003). "Some Causal Lessons from Macroeconomics." *Journal of Econometrics* 112: 121-125.

——. (2005). "Automatic Inference of the Contemporaneous Causal Order of a System of Equations." *Econometric Theory* 21: 69-77.

Lewis, David. (1973). "Causation." *The Journal of Philosophy* 70 (17): 556-567.

——. (1979). "Counterfactual Dependence and the Direction of Time's Arrow." *Noûs* 13 (4): 455-476.

Mackie, J. L. (1974). *The Cement of the Universe: A Study of Causation*. Oxford: Clarendon.

Menzies, Peter. (2008). "Counterfactual Theories of Causation." *The Stanford Encyclopedia of Philosophy*. Retrieved March 30, 2009, from http://plato.stanford.edu/entries/causation-counterfactuals/

Pearl, Judea. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Plato. (1971). *Meno*. Trans. W. K. C. Guthrie. New York: Bobbs-Merrill Company, Inc.

Reiss, Julian. (2008). *Error in Economics: Towards a More Evidence-Based Methodology*. New York: Routledge.

Rosenzweig, Mark R., and Kenneth I. Wolpin. (2000). "Natural 'Natural Experiments' in Economics." *Journal of Economic Literature* 38 (4): 827-874.

Scheines, Richard. (2005). "The Similarity of Causal Inference in Experimental and Non-experimental studies." *Philosophy of Science* 72: 927-940.

Shipley, Bill. (2000). *Cause and Correlation in Biology*: *A User's Guide to Path Analysis, Structural Equations and Causal Inference.* Cambridge: Cambridge University Press.

Woodward, James. (2008). "Causation and Manipulability." *The Stanford Encyclopedia of Philosophy.* Retrieved March 30, 2009 from http://plato.stanford.edu/entries/causation-mani/