Finding front doors

Christoph Hanck

Summer 2024

Motivation

- Modelling the DGP, listing all the paths, finding a set of variables that close all the back doors, and measuring and controlling for all of those variables is a lot of work.
- Controlling for everything is difficult, especially in social sciences where many things might matter and you might run into a variable you have to control for but you cannot.
- There are so many variables that might be on back doors, sometimes there are high chances to forget some of them in your causal diagram.
- An alternative approach the answer a research question is to, instead of actively closing back doors, find ways of isolating just front doors.

How to estimate the front doors directly?

The front door method

We can estimate individual arrows on the front door paths even if the overall effect is not identified.

Example: Wealth on Lifespan

- You want to estimate the effect of Wealth on Lifespan
- There are many back doors where we do not have data, for example "business skill"
- However, since people buying the lottery got their wealth from many sources like inheritance, working, etc., these sources are not related to *whether you win the lottery*
- So we may not consider these as backdoors on variation

Trying to push a string

How can we pick out just the variation we want?

- It all comes down to the fact that our *treatment variable varies for different reasons.*
- The key idea here is that we can partition the *variation in treatment*.

  We can select a particular sample or use certain statistical adjustment to throw out the part that is driven by back doors.
- The cleanest application of this approach is the **randomised controlled experiment.**

Randomised controlled experiments

Definition: Randomised controlled experiment

In a **randomised controlled experiment**, the researcher assigns treatment (or the absence of treatment) to people, and watches the resulting differences in outcome.

Randomised controlled experiments

Why do they work?

- Experiments work because they create a form of *variation in the treatment* that has no back doors.
- If the treatment was assigned randomly, then for everyone in the experiment, variation in all the variables on all the back doors should be unrelated to whether they got the treatment or not.
- Hence, all the back doors will be closed.

Randomised controlled experiments

Example: Charter school's influence on grades

- We are interested in figuring out whether charter schools improve students’ test scores more than traditional public schools.
- Many variables influence people's decision to attend Charter school, e.g. background, personality, location, etc.

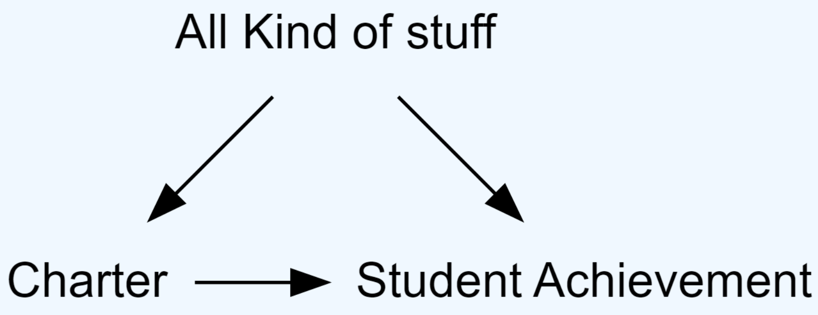

Figure 1: Causal diagram of Influence of Charter School on Grades

Randomised controlled experiments

Example: Charter school's influence on grades

- Many students in charter schools are assigned slots based on a lottery system.
- There are no back doors from Lottery to *Student Achievement*.
- The effect of *Lottery Win* on *Student Achievement* is identified in the data without any controls.
- The only way that Lottery can affect *Student Achievement* is through *Charter*. So if the effect of *Lottery Win* on *Student Achievement* is calculated, then that can give us a hint about the effect of *Charter* on *Student Achievement*.

Randomised controlled experiments

Example: Charter school's influence on grades

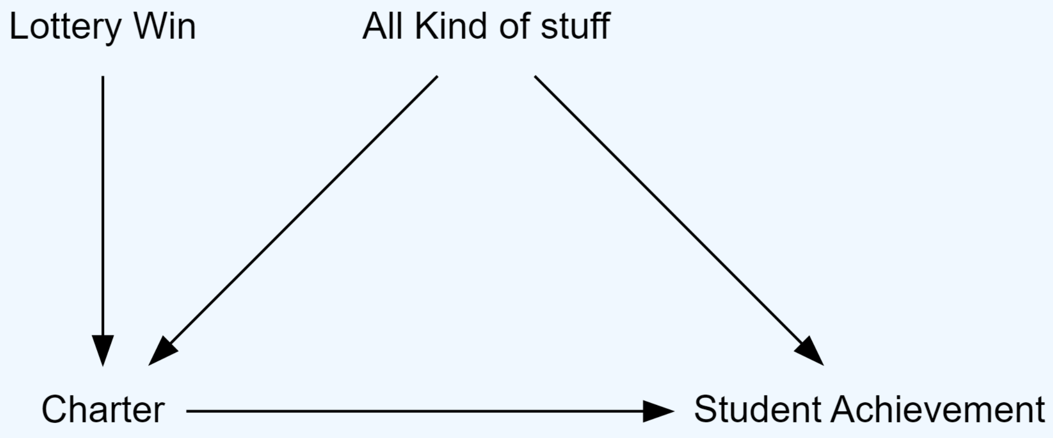

Figure 2: Causal diagram of influence of Charter schools on grades

Randomised controlled experiments

Example: Charter school's influnece on grades

The data from the experiment can be analysed in two ways:

1. - Throw out all the data where Charter is not driven by Lottery, then look at the effect of *Charter* on *Student Achievement*.
   - Take only the students who were in the lottery and compare the ones who got into the Charter against the ones who did not.
2. - Use *Lottery Win* to explain or predict *Charter*. Then, take the prediction of whether someone goes to a *Charter*, which is based purely on their *Lottery Win*.
   - Then look at the relationship between the prediction and *Student Achievement* to get the effect.

Natural experiments

Definition: Natural experiment

Natural experiments are a real-world setting in which randomisation of treatment and control conditions has been already done by nature or by other factors that are outside our control

Natural experiments work because they fix some of the variation in treatment to *have no back doors*.

Example: Charter school's influnece on grades

In the last example the charter schools did some randomisation to select the pupils—the randomisation occurred in the world and we could take advantage of it.

Source of exogenous variation

Definition: Source of exogenous variation

A source of variation in the treatment that has no open back doors is called a **source of exogenous variation**.

An ideal source of exogenous variation is not caused by any other variable that belongs on the causal diagram.

Sources of exogenous variation

Anything which is random in the context of our DGP can be considered as a source of exogenous variation.

The causal diagram for this looks almost exactly like a **randomised controlled trial.**

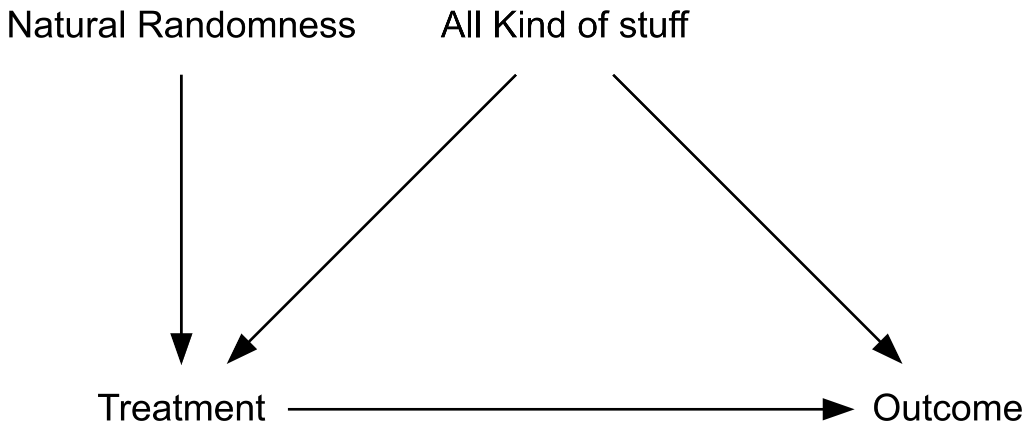

Figure 3: Causal Diagram for Exogenous Source of Variation

Sources of exogenous variation

Randomised controlled experiments vs. natural experiments

- Natural randomness can have *back doors to the outcome* but pure randomisation does not.
- Natural experiments are *more natural* than randomised controlled experiments. People may not even realize that they are a part of an experiment.
- The observations from natural experiments tend to be more realistic than randomised controlled experiments.

Sources of exogenous variation

Randomised controlled experiments vs. natural experiments.

- Sample sizes are *bigger* in natural experiments than in randomised controlled experiments.
- We *isolate the variation in treatment* that is driven by the natural randomness. So we are seeing the effect only among people who are sensitive to NaturalRandomness.

  No such thing happens in Randomised Controlled Experiments!
- People believe more in **exogeneity** of pure randomisation than natural randomisation.

Sources of exogenous variation

Example: the effect of air pollution on number of people driving

- **Pan He** and **Cheng Xu** in 2019 look at whether **air pollution being worse causes people to drive more**.
- He and Xu get their data in Beijing. They look at **whether people drive more on days when there is more pollution**, and find that **they do**.
- Problem: there's also obvious reverse causality issues as well as back doors (factories running, etc.)
- They find an **exogenous source of pollution variation in the direction of the wind** - In Beijing, a west-blowing wind blows pollution into the city.

Sources of exogenous variation

Example: the effect of air pollution on number of people driving

- By isolating just the variation in pollution driven by wind direction, they find that an increase in daily pollution large enough to change the government's rating from "not polluted" to "polluted" increases driving by 3%.
- There can be **back doors**: for example, the **direction of the wind** might change with the season, and the season is related to pollution and driving.

Example: the effect of air pollution on number of people driving

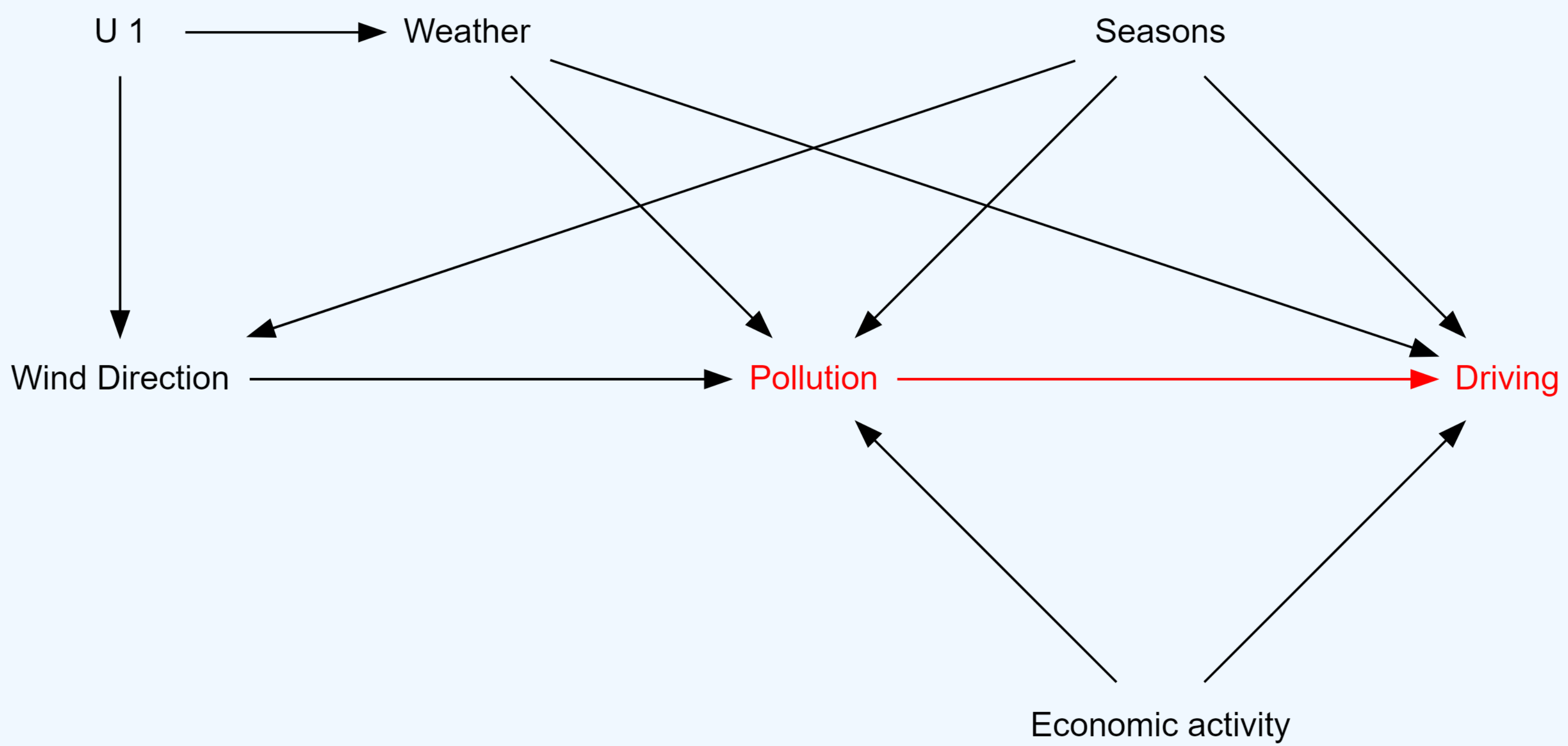

Figure 4: Causal Diagram of effect of Air Pollution on number of people driving

Sources of exogenous variation

Example: the effect of uncompensated health care on patient experience

- Camilleri and Diebold (2019) looked at the effect of uncompensated care (medical care given by hospitals that do not get paid for) on patient experience.
- The amount of uncompensated care a hospital gives is related to some **back-doors**:
  - **The kinds of patients they get**
  - **How likely those patients are to pay**, etc
- Camilleri and Diebold use the 2014 Medicaid expansion as a source of exogenous variation, in which only some states **expanded access to the Medicaid program.**
- Using this source of exogenous variation, they find that **reductions in uncompensated care did not improve patient experience much**.

Sources of exogenous variation

Example: the effect of uncompensated health care on patient experience

- The choice of states to accept Medicaid was not random, but rather **politicised**.
- Medication expansion may **not be a source of exogenous variation**.
- Medicaid expansion, and thus **expanded access to insurance**, should change lots of things about health care besides hospital compensation that might also be related to patient experience.
- We have to assume that the e**xpansion only affected hospital compensation** in order to consider **variation in compensation driven by the Medicaid expansion as being exogenous**.
- We need to be very careful in thinking what assumptions we have to make about the data generating process, and whether those assumptions are true.

More on randomisation

- **Isolating front paths is always feasible**, just like identifying the effect of a treatment by closing back doors is always feasible, even if we do not have anything even remotely like purely-random variation as we would in a randomised experiment or even a lottery.
- However, the farther away we get from that pure randomisation, the more things we need to control for, and the more assumptions we have to make, and perhaps the more unbelievable assumptions we have to make.
- We are merely replacing the difficulty of finding and closing all back doors for a treatment variable with the difficulty of finding and closing all back doors for something else.

Identifying a causal effect

Example: Front door method

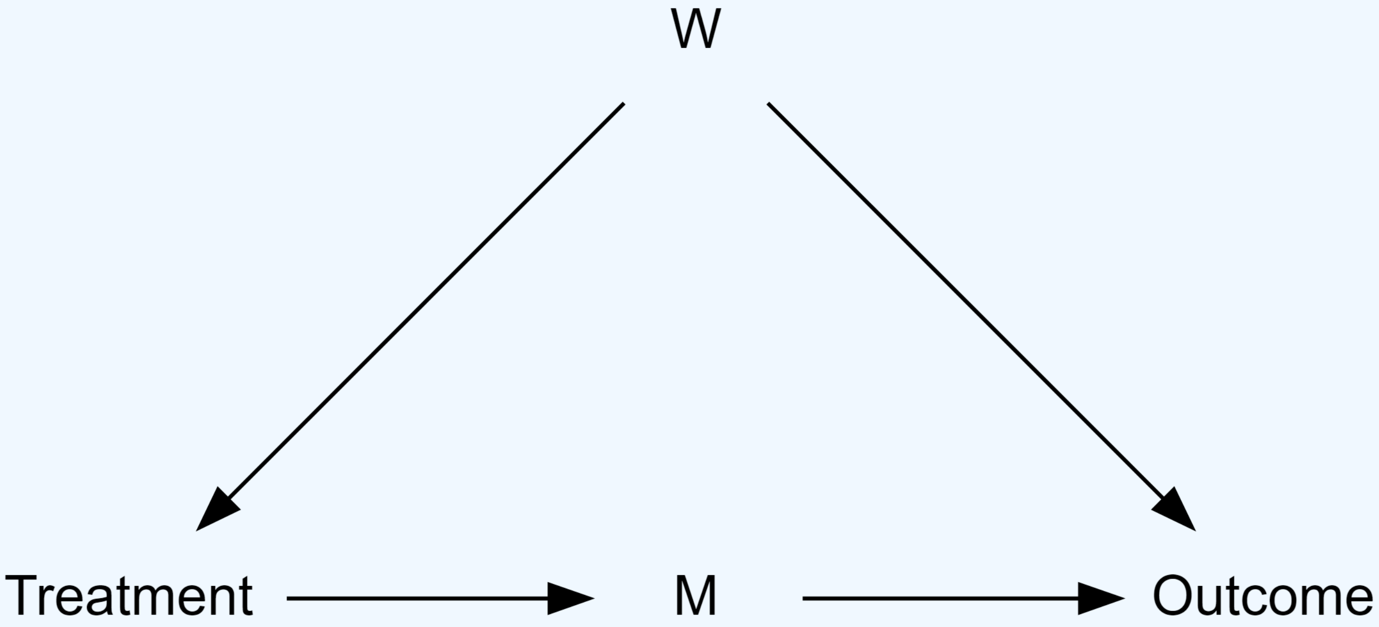

Figure 5: Causal Diagram when using the front door method

Identifying a causal effect

- There is another way to identify the causal effect of a treatment on an outcome by isolating *front doors*:

  It is called, appropriately, the **front door method**.

- The front door method works when your causal diagram looks something like Figure 5, when there is a bad path that cannot be closed, such as if W in that diagram cannot be measured.
- If W cannot be measured, then we cannot control for anything to identify the effect of Treatment on Outcome.

Identifying a causal effect

Example: Front door method – ctd.

We can identify that the only back doors are:

- Treatment ← W → Outcome ← M
- M ← Treatment ← W → Outcome

We can identify Treatment → M and M → Outcome

We just need to combine those two effects to get our effect: Treatment → M → Outcome.

Identifying a causal effect

Example: Smoking causes cancer

- By Pearl and Mackenzie 2018
- It is difficult to figure out the effect of smoking on something like cancer rates because there are lots of things related to whether you smoke (background, income, health-mindedness, etc.)
- There are a lot of back doors that cannot be closed.
- Let’s say that there is something called *TarInLungs* that sits between smoking and cancer. In this simplified fantasy, the only reason *Smoking* causes *Cancer* is because it causes *TarInLungs*, and *TarInLungs* causes *Cancer*.

Identifying a causal effect

Example: Smoking causes cancer

The causal diagram will look like:

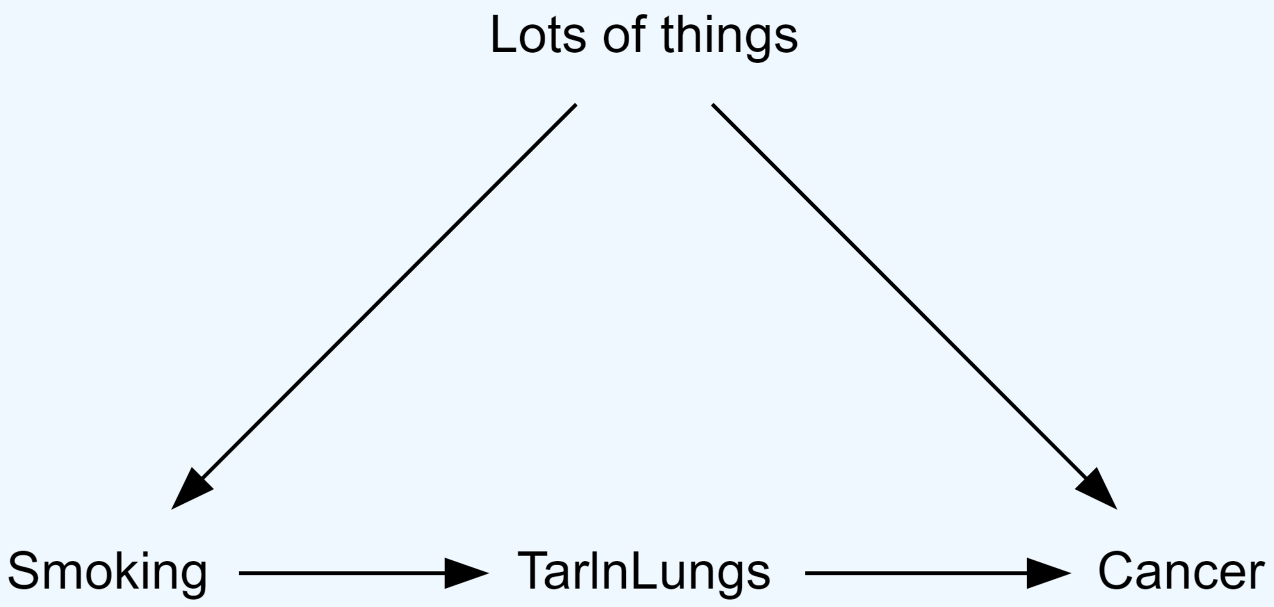

Figure 6: Causal Diagram for Smoking and Cancer

Identifying a causal effect

Example: Smoking causes cancer

- Given this diagram, let's say that we look at the raw, unadjusted relationship between *Smoking* and *TarInLungs* and find that an additional cigarette per day adds an additional 15 grams of tar to your lungs over 10 years.
- Then, we look at the relationship between *TarInLungs* and *Cancer* while controlling for *Smoking* and find that an additional 15 grams of tar in your lungs increases the chances of getting cancer by 2% over your lifetime.
- So, then an additional cigarette per day increases the tar in your lungs by 15 grams, which in turn increases your probability of cancer by 2%. So an additional cigarette per day increases your probability of cancer by 2%.
- That is the front door method!