

First Partial Exam: Notes

Econometrics II

Max Heinze (mheinze@wu.ac.at) Sannah Tijani (stijani@wu.ac.at)

Department of Economics, WU Vienna

Department of Economics, WU Vienna

December 4, 2025

Questions on Statistical Learning

In **Module 1**, we said:

Different methods have different degrees of **flexibility**, i.e. they can only produce a narrow range of possible shapes of f . For example, linear regression is rather inflexible.

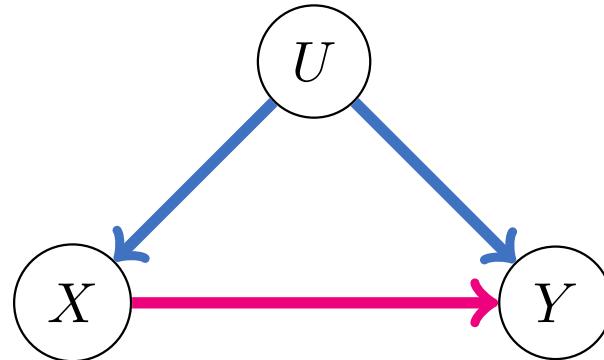
The **benefit** of choosing a flexible approach is evident. However, there is an important downside: **More flexible** methods yield results that are **less easy to interpret**.

Most of you answered the question by saying **unsupervised learning** is most suited for prediction. Actually, I wanted to hear the more general statement that a **flexible model** should be used. Unsupervised learning is certainly a very flexible approach, but there are other flexible models too. The main point of the question was to ask about the **interpretability-flexibility tradeoff**.

In the exam, no points were deducted for mixing up the two since this answer was extremely common.

Confounders

Confounders are variables that influence both the explanatory and the dependent variable:

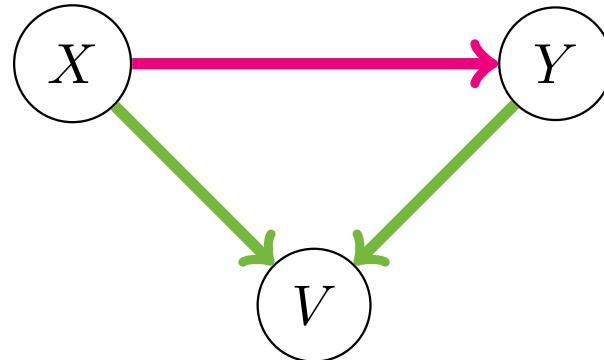


Since a confounder must **lie on a backdoor path**, it cannot be either the explanatory or the dependent variable itself.

In the exam, there was **one confounder**, and there were **two backdoor paths** running through that confounder (same in all variants of the question).

Colliders

Colliders are variables that are causally influenced by at least two variables, no matter which ones:



This definition is different from the original definition we used in class. The original definition was not perfectly accurate. Please update your notes and remember this definition.

This means that colliders can be anywhere in the DAG, including the explanatory and dependent variables themselves.

In the exam, we accepted: **no colliders** (old definition), **three colliders** (this definition), **two colliders** (this definition minus the dependent variable of interest).

Simultaneity vs. Reverse Causality

Remember the definitions we had for **simultaneity** and **reverse causality**:

- **Reverse causality** is when $X \leftarrow Y$
- **Simultaneity** is when $X \leftrightarrow Y$

If we use those definitions, the **two issues cannot appear simultaneously**.

- With **pure reverse causality**, the issue is determining the direction of causation.
- If we have **simultaneity** we need to disentangle the two effects.

Reading and Understanding

We noticed that some of you had trouble understanding the exam questions. One case was the word “**demographics**.”

If you do not understand a word or a question, ask.

Some of you did not understand, *asked* during the exam, and got full points on the question. Others did not and lost points because they later said demographic variables were missing.

Also, if you do not know the answer, don’t start making something up that, at best, broadly matches the vibe of the question. It likely won’t be the correct answer.

We can notice the difference between somebody who knows the answer and somebody who knows three fancy words that should appear there (confounder, simultaneity, endogenous, ...) but does not know why.