

An Empiricist's Checklist[†]

Erkmen Giray Aslim

There are four important questions that need attention in an empirical study. Angrist and Pischke (2008) refer to these questions as frequently asked questions (FAQs) due to its fundamental role in econometrics. These questions are defined as follows:

1. What is the causal relationship that is of concern for the researcher?
2. What is the experimental design that allows the researcher to capture the causal effect?¹
3. What is the identification strategy?²
4. What is the procedure of statistical inference?

The answers to four questions defined above are sometimes ill-specified in research projects.³ For example, testing a specific hypothesis and/or the data used in the process might have limitations. Then the answers to these questions are not straight-forward and often require rigorous treatment. Since the issues above constitute the building blocks of empirical research, it is a worthy checklist for the researcher.

Unlike certain fields in social sciences, economics deals with models of cause and effect. Needless to say, causality requires stronger assumptions than simple associations. There might be cases in which the identification strategy is jeopardized while in search for a causal effect. If the independent variable is correlated with the error term, then we no longer have a causal relationship (i.e. $E[u|X] \neq 0$). Card (2001) reviews the set of

[†]One shall consider this note as a summary of Angrist and Pischke's (2008) first chapter.

¹This question can be asked in many different forms, as preferred in economics textbooks, with a direct approach of introducing the model.

²The identification strategy allows the researcher to disentangle the relationship between the variables of interest. Many studies in econometrics deal with point identification. This method does not allow for uncertainty when the values of the object, so-called estimates, are inferred from the model.

³The issue mostly arises on the basis of causality. One is more likely to cross the red line, for example, when interpreting the causal link between returns to education and annual salary with the absence of "innate ability".

studies on returns to schooling that tackle with the causality issue by using instrumental variables. He shows that the literature is converging to a point that institutional factors in education is used as a credible instrument to factor out the effect of education on wages. Instrumental variables should be chosen with caution because the results may be sensitive to the specific instrument.

The methodology of the experiment in determining the causal effect can be highly controversial but the role in supporting the hypothesis is always favorable. This controversy was obvious in the experimental study of Milgram (1963).⁴ The lesson from this experiment for an empiricist is to think about the correct experimental design to answer the research question in mind. Then there might be room for policies. Some research questions, however, cannot be answered by an experiment. The effects of start age on first grade scores, for example, cannot be identified. The discussion related to the issue is relatively simple and can be supported by two facts. If we are comparing age 6 and 7, then by the maturation effect older kids are going to get better test scores. Suppose we fix the age by allowing the 6 year old kids to finish first grade and take the test with those who start at age 7. Because 6 year old kids spend more time in school, a year, they will be more successful if school actually contributes to the learning process.

The identification strategy is based on deducing the unknown value of the parameter from the distribution of the observational data.⁵ Identification can be challenging under the presence of omitted variables (endogeneity can also be considered here), misspecification issues, measurement errors in regressors, and a nonrepresentative data. After we deal with all these issues, the mode of statistical inference gains importance. The researcher should provide, for example, a justification for using homoskedastic standard errors. Statistical inference can also be challenging when using clustered or grouped data. Microeconometrics gives us certain set of tools to tackle these challenges. However, there might still be issues related to the model after we satisfy our checklist. I leave the discussion of those issues for a later blog entry.

⁴He convinced the subjects in the experiment to give electric shocks to protestors.

⁵Observational data is obtained by sampling the relevant population of subjects without controlling any characteristics (Cameron and Trivedi, 2005).

References

- [1] Angrist D. J., J. Pischke, 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- [2] Card, D., 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica*, 69(5), 1127-1160.
- [3] Cameron, A. C., P. K., Trivedi, 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press.
- [4] Milgram, S., 1963. "Behavioral Study of Obedience." *Journal of Abnormal and Social Psychology*, 67, 371-378.