Section 5: Validity Threats and Measurement Bias

1 Introduction

Throughout this course we've developed several types of models. We have discussed different regression strategies to tease out the causal effect of potential variables of interest. Today we will be discussing from a practical standpoint what might threaten the validity of our model – from the assumptions our regression is built on to limitations in the underlying data.

Q1: What are some factors that could hurt the validity and usefulness of our model?

2 Internal and External Validity

Internal Validity: Have we estimated valid causal effects for the population under study? Do our confidence intervals have the correct coverage rates?

External Validity: Can we generalize the results to other populations and settings?

2.1 Threats to Internal Validity

Q2: What do (most) threats to internal validity have in common?

Below we have listed some potential threats to internal validity. Please explain why you think these factors are or are not threats to internal validity.

Q3.1: Sample selection bias

Q3.2: Omitted variable bias

Q3.3: Simultaneous causality when RHS variable is measured at a later time than the outcome variable on the LHS

Q3.4: Perfect multicollinearity

Q3.5: Simultaneous causality when RHS variable is measured at an earlier time than the outcome variable on the LHS

Q3.6: Wrong Functional Form

Q3.7: Errors-in-variables bias

Q3.8: Bad controls

Q3.9: Low/High R-squared

Q4: What threat to internal validity does not violate the assumption of conditional mean independence? How does it affect our model?

2.2 Threats to External Validity

Consider the paper, "Sensation Seeking, Overconfidence, and Trading Activity." Authors Mark Grinblatt and Matti Keloharju attempt to look at the psychological factors of overconfidence and sensation seeking as influential to trading behavior. They used speeding tickets as a proxy for sensation seeking and mandatory psychological profiles to characterize overconfidence.

This paper consolidated several different data sets to build a coherent analysis. They used trading data of household investors domiciled in Finland from 1995 to 2002. Additionally they used HEX stock data, data from the Finnish Vehicles Administration, and the psychological profiles of Finnish soldiers (It should be noted that military service is compulsory for men in Finland).

Q5: Can you think of any external validity concerns with these data sets?

Q6: Can you imagine any internal validity concerns? Would it help to know more about Finland, the methods used, and the data sets themselves?

2.3 Errors-in-Variables Bias (Measurement Error)

We have discussed three types of measurement error:

- Measurement Error in X
- Measurement Error in Y
- Measurement Error in X with Controls

Q7: What do we mean by 'Measurement Error in X'? What does this look like in practice?

Q8: What is this type of bias called?

Q9: How does this bias (Measurement Error in X) our models?

Q10: How is measurement error in Y different than attenuation bias?

Q11: True or False: Adding controls helps us eliminate attenuation bias.

2.4 Bad Controls

Q12: Why do we add controls?

Q13: Intuitively what makes a control variable bad?