KKV Chapter 5 “Understanding What to Avoid”

Sweeping Chapter Overview:

KKV are concerned with understanding sources of inefficiency and bias and reducing inefficiency and bias to “manageable proportions” (p. 150) by controlling the research. This process occurs after we have selected our observations.

Definitions:

Unbiasedness: Our inference is the right estimate.

Efficiency: Narrowing the interval around the right estimate.

The High Points of Reducing Bias and Inefficiency:

5.1 Measurement Error

- can bias results and make them less efficient
- measure data as nominal, ordinal, or interval based on what is most appropriate for your theoretical purposes (e.g. don’t assign a numerical ranking to the religion to which a person belongs)
- The effects of systematic vs. non-systematic measurement error:
    1. Systematic m.e. (consistently overestimating certain units) causes bias in descriptive inference while only certain types of systematic m.e. (systematic m.e. that not do affect all units consistently) will bias causal inferences. Avoid systematic m.e by using measurement judgments made for different purposes by other researchers.
    2. Non-systematic (random) m.e.: “Error in the dependent variable causes inefficiencies [but no bias], which are likely to produce incorrect results in any one instance and make it difficult to find persistent evidence of systematic effects” (p. 155). Non-systematic m.e. in the key causal variable biases inferences (a weaker than actual causal relationship will be shown).

        The implications for non-systematic m.e. in the key causal variable leads KKV to two conclusions (cited directly from p. 165-66):

        1) If an analysis suggests no effect to begin with, then the true effect is difficult to ascertain since the direction of bias is unknown; the analysis will then be largely indeterminate and should be described as such. The true effect may be zero, negative, or positive, and nothing in the data will provide an indication of which it is.
        2) However, if an analysis suggests that the explanatory variable with random measurement error has a small positive effect, then we should use the results in this section as justification for concluding that the true effect is probably even larger than we found. Similarly, if we find a small negative effect, the results in this section can be used as evidence that the true effect is probably an even larger negative relationship.

5.2 Excluding Relevant Variables: Bias

KKV discuss **Omitted Variable Bias**.

Rule 1: Irrelevant o.v. cause no bias.

Rule 2: If the o.v. is uncorrelated with the included explanatory variable, there will be no bias.

Rule 3: Given Rules 1 & 2, we can omit control variables, even if the have a causal relationship with our dependent variable, as long as they are uncorrelated with our explanatory variable.

Note: While we have no bias in our causal inference, we do lose accuracy in forecasting values of the dependent variable.

Rule 4: Control for o.v. that we think are biasing our results (or if that is not possible – because we don't have a good source of data – ascertain the direction of the bias).

Rule 5: Do not control for an explanatory variable that is, in part, a consequence of our causal variable.

5.3 Including Irrelevant Variables: Inefficiency

If you control for an irrelevant explanatory variable, you will still get an unbiased estimate. If the irrelevant e.v is uncorrelated with the key e.v., the efficiency of our estimate remains that same as if we had not included the irrelevant variable.

If that is not possible: Use our a priori knowledge of the available observations and select observations and assign the values of the explanatory variables to avoid bias and inefficiency.

In any research report you should make clear your assignment and selection process. Discuss (1) rules used and (2) hidden sources of bias and what, if anything, you did about them.

5.6 Controlling the Research Situation

In small-n studies, control by matching. Searching for "'unusual' observations in order to get a close match across groups" (p. 204) risks omitted variable bias. One productive approach for dealing with this problem is to choose case studies by matching but observations within those case studies by other criteria.