**Formulation of a thesis question and empirical approach suggestion,** Sofia Villas-Boas

The way to start the honors process is to find a faculty adviser or a ARE graduate PhD student adviser (mentor) that helps you narrow down topic and works with you: <https://are.berkeley.edu/people/faculty>, <https://are.berkeley.edu/people/students>.

The link to the honors program information with timelines (<https://nature.berkeley.edu/advising/honors-program>) after you start the two semester process is here <https://are.berkeley.edu/eep/undergraduate-program>

Some advisers like you to approach them with a question already very well formulated and the ideal data set and policy variation identified (data sources found and background leg work done on how the data source has the data variation needed to identify the effect of interest)

What is the marginal effect of x1 on Y?

this is important because bla bla bla

I am concerns that in answering the question x1 is not randomly assigned and the analysis of y=alpha + x1 beta1+ epsilon

could suffer from omitted variable bias

because of that I will also use x2, x3 etc in the analysis to control for the omitted variable problem and investigate the conditional average marginal effect of x1 on y controlling for x2 and x3

y=alpha + x1 beta1 + x2 beta2 + x3 beta3 + epsilon

When I do that I notice that the estimated beta1 hat changes as I expected, as the estimate of the underspecified model had a positive/ negative bias

In addition I also have data on y x1, x2 and x3 for many observations i and also for the same observations i over T periods in time. I have thus a panel data set.

With that additional data I can estimate  the model by including individual fixed effects, and also time fixed effects.

In so doing I control for bla bla bla that could also bias the estimate of beta1

If x1 is a policy and the policy was implemented around a cut off you can do an RD design

If x1 is a policy implemented in a  subset of individual observations in a certain period in time you can do a difference in differences with fixed effects also