ARE213 Problem Set #1B

Peter Alstone & Frank Proulx

October 16, 2013

1 Problem #1

1.1 Part A

Under the assumption of random assignment conditional on the observables, what are the sources of misspecification bias in the estimates generated by the linear model estimated in Problem Set 1a?

Wrong functional form. In Problem Set 1A we used linear (i.e., $y = \beta x + \epsilon$) estimators to make "corrections" while the true functional form of the relationships between the covariates we included in the modern were certainly not linear. By imposing a linear function on a non-linear data generating process (described by the CEF), we introduce misspecification bias in the model.

Omitted Variables Bias. We were able to use variables included in the dataset in our linear model, but not the unobserved variables that may be important for control. If omitted variables exist that both determine outcomes related to birth weight and are correlated with smoking status we will over- or under-estimate the effect (depending on the characteristics of the omission).

1.2 Part B

Now, consider a series estimator. Estimate the smoking effects using a flexible functional form for the control variables (e.g., higher order terms and interactions). What are the benefits and drawbacks to this approach?

We can attempt to reduce the magnitude of the first source of bias mentioned above (Functional Form) by introducing non-parametric series estimators as a replacement for linear regression.

2 Problem #2

The Propensity Score Method (PSM) uses a "surrogate" normalized metric (p-score) as a replacement for the observable controls that would normally be used to condition the estimates of a treatment response to the variable in question.