PROBLEM SET II

1. 2SLS is a many-splendored thing

Consider the simple 2SLS model we discussed in class (see also MHE 4.6.4):

$$y_i = x_i'\beta + \eta_i$$
$$x_i = z_i'\pi + \xi_i$$

where x_i is a scalar and there are Q instruments.

a. Show that IV using estimated first-stage fitted values (from an OLS first stage) as an instrument for x_i is the same as 2SLS (i.e., OLS with estimated first-stage fitted values substituted for x_i).

b. Now, suppose you know the first stage coefficient, π, and use this to construct "known fits,"

$$x_i^* = z_i'\pi$$

Show that that the IV estimator that uses x_i^* as an instrument for x_i has the same limiting distribution as IV using OLS-estimated fits and therefore IV using a known first stage to construct instruments has the same limiting distribution as conventional 2SLS.

c. (extra credit) Show that manual 2SLS using the known first-stage fitted values (i.e., substituting x_i^* for x_i) does not have the same limiting distribution as regular 2SLS. Why not? Can you say which is more efficient, manual 2SLS with known fits or the conventional 2SLS estimator in part b?

2. A little LATE

a. Show that if the first stage for a Bernoulli instrument comes from heterogeneous potential assignments:

$$D_i = D_{0i} + (D_{1i}-D_{0i})Z_i,$$

but treatment effects are constant so that:

$$Y_{1i}=Y_{0i}+\rho,$$

then the IV/Wald estimand is ρ, with or without monotonicity. Derive a formula for the "bias" induced by failures of monotonicity in the heterogeneous effects model and use this to explain why failures of monotonicity need not be fatal. (Here, bias means the difference between LATE and the IV estimand without monotonicity.)

b. Show that in a world with heterogeneous effects, monotonicity is necessary for sensible IV, in the following sense: without it, $Y_{1i}-Y_{0i}$ might be positive for everyone and yet the reduced form equal to zero or even negative.

c. Show that for instruments that generate no always-takers, LATE is TOT, while for instruments with no never-takers, LATE is the effect of treatment on the non-treated (TNT).

3. Prove that the marginal means of any function of potential outcomes, Y_{1i} and Y_{0i}, are identified. Use this to show how to estimate the effect of treatment on distributions for compliers.

4. Structure your thinking

Discuss the link between causal effects and structural parameters in a Bivariate Probit model of the relationship between divorce and female labor force participation (See MHE 4.6.3). The purpose of the model is to determine whether female employment strengthens a marriage or increases divorce. Organize your discussion as outlined below:

        a. Explain why the causal effect of employment on divorce is difficult to determine. Can you design an experiment to answer the question of interest?

        b. Write the potential outcomes and potential treatment assignments in your causal model in terms of latent indices with unobserved random errors in a structural model.

        c. What should the population be for this study? What does it mean for employment to be "endogenous" in the structural model? How about in the causal model?

        d. Show how to use the Probit structural parameters and distributional assumptions to calculate the population average treatment effect (ATE), the effect on the treated (ETT), and LATE. Discuss the likely relationship between these average causal effects.

        e. Using data the Angrist (1999) LDV paper to estimate and compare:

        i.   OLS estimates to Probit MFX
        ii.   IV estimates to Bivariate Probit MFX

5. Diffs-in-diffs

a. Prove that with two periods, fixed effects estimation using micro panel data is the same as estimation in first-differences.

b. Card (1992) estimates the effect of federal minimum wage using a two-period state panel of 102 observations, exploiting the fact that in rich states the federal min is irrelevant, while in poor states it's a big deal. Derive Card's estimator from micro data (i.e., start with a model for micro data and explain why estimating Card's grouped equation makes sense in this context). Finally, propose an alternative possibly more efficient procedure that also uses state-by-year means.

6. Regression Discontinuity Designs

a. Use the Lee (2008) data (posted in the MHE data archive) and the IK software to produce IK (i.e., nonparametric RD) estimates of the effects of incumbency on election victories. Compare these to parametric estimates.

b. Check your IK estimates by taking the bandwidth as given and using the appropriate weighted OLS procedure in Stata.

14.387 Applied Econometrics: Mostly Harmless Big Data
 Fall 2014