MIT 14.661 Angrist, Pischke, Menzel
Fall 2006-2007
Due: October 19, 2006

GRADUATE LABOR ECONOMICS
PROBLEM SET II

IV estimates of earnings functions

You are interested in estimating a regression of log wages, y_i, on years of schooling, s_i, while controlling for another variable related to schooling and earnings that we will call a_i.

1. Consider the following regression equation:

$$y_i = \beta + \rho s_i + a_i\gamma + \varepsilon_i \quad (1)$$

Assume that the regression coefficients β, ρ and γ are defined in equation (1) such that ε_i is uncorrelated with s_i and a_i.

a. Suppose you estimate a bivariate regression of y_i on s_i instead. What is the plim of the coefficient on s_i in terms of the parameters in equation (1)? When does the "short regression" estimate of ρ equal the "long regression" estimate?

b. Why is the long regression more likely to have a causal interpretation? Or is it?

2. Consider using information on quarter of birth, Q_i (= 1, 2, 3, 4), as an instrument for equation (1) when a_i is unobserved. You are trying to use an instrument to get the long-regression ρ in a sample of men born (say) in 1930-39.

a. What is the rationale for using Q_i as an instrument?

b. Show that using $z_i = 1[Q_i=1]$ plus a constant as an instrument for a bivariate regression of y_i on s_i produces a "Wald estimate" of ρ based on comparisons by quarter of birth. Given the rationale in (a), is this estimator consistent for ρ in equation (1)?

c. Suppose that the omitted variable of interest, a_i, is still unobserved but we know that it is the age of i measured in quarters. What is the plim of the Wald estimator in this case? Can you sign the bias of the Wald estimator?

d. Suppose that instead of using z_i, you use Q_i itself as an instrument. Show that the resulting estimator is not consistent either (continuing to assume a_i is omitted and equal to age in quarters). Can you use the two inconsistent estimators (Wald and IV using Q_i) to produce a consistent estimate of ρ?

e. Now suppose that a_i is observed. Explain how you can consistently estimate ρ by 2SLS using 3 quarter of birth dummies, $z_{1i} = 1[Q_i=1]$, $z_{2i} = 1[Q_i=2]$, and $z_{3i} = 1[Q_i=3]$, plus a constant as the excluded instruments. Give necessary and sufficient conditions for this to work.

f. As an alternative to 2SLS, consider using a dummy for “middle quarters”

$z_{im} = I[Q_i=2 \text{ or } Q_i=3]$

plus a constant as instruments for a bivariate regression of y_i on s_i. Show that this also produces a consistent estimate of ρ when Q_i is uniformly distributed (still assuming that a_i is age in quarters). Explain why this strategy works. On what basis might you choose between these alternative estimators?

f. Suppose the equation of interest includes a quadratic function of age in quarters:

$$y_i = \beta + \rho s_i + a_i\gamma_0 + a_i^2\gamma_1 + \varepsilon_i \qquad (2)$$

Explain why the "middle-quarters" estimator no longer works. Can you think of an estimator that does?

3.Construct an extract from the 1980 Census similar to the one used by Angrist and Krueger (1991). use this extract to compute and compare the estimates discussed in questions 1 and 2. Hint: Get your data from the IPUMS.