Midterm

Economics 435: Quantitative Methods

Fall 2010

Please answer the question I am asking - no more and no less - and remember that the correct answer is usually short.

1 Cluster samples

Cluster samples are built by randomly selecting clusters or groups of individuals, and then randomly sampling individuals within each group. They have many of the same properties as random samples, but there are also some important differences.

Suppose we have sampled $n > 1$ randomly selected students, indexed $i = 1, 2, \ldots, n$ in each of $S > 1$ randomly selected schools, indexed $s = 1, 2, \ldots, S$. Our total sample size is nS. For each individual we observe some random variable x_{is}. Without loss of generality, let:

$$x_{is} = \mu + v_s + u_{is}$$

where μ is an unknown constant, and v_s and u_{is} are unobserved random variables such that:

$$
\begin{aligned}
E(v_s) &= 0 \\
E(u_{is}|v_s) &= 0 \\
cov(u_{is}, u_{js}) &= 0 \qquad \text{for all } j \neq i
\end{aligned}
$$

That is, the overall mean of x is μ, the mean in school s is $\mu + v_s$, and u_{is} captures any remaining heterogeneity across individuals. Let $\sigma_v^2 = var(v_s) > 0$ and let $\sigma_u^2 = var(u_{is}) > 0$.

a) Let $\bar{x}$ be the sample average:

$$\bar{x} = \frac{1}{nS} \sum_{s=1}^{S} \sum_{i=1}^{n} x_{is}$$

Is $\bar{x}$ an unbiased estimator of μ? Prove it.

b) Find $var(x_{is})$.

c) Find $corr(x_{is}, x_{js})$ for $i \neq j$.

d) Find $var(\bar{x})$.

e) Under these assumptions, is the variance of $\bar{x}$ in this cluster sample always higher than that of a traditional random sample of the same size?

f) The asymptotic properties of a cluster-based sample depend on whether we think of the number of clusters going to infinity or the number of observations per cluster going to infinity. Keeping n fixed, is it the case that

$$\text{plim}_{S \to \infty} \bar{x} = \mu$$

g) Keeping S fixed, is it the case that

$$\text{plim}_{n \to \infty} \bar{x} = \mu$$

2   Permanent income and the black-white test score gap

This question is based on a recent working paper by Rothstein and Wozny (available at `http://www.econ.sfu.ca/Seminars/SeminarPapers/rothstein_wozny_2ndpaper.pdf`) by the same title. I have taken liberties with a few details of the paper to make a tractable exam question.

Researchers in the economics of education have investigated various explanations for the large and persistent gap in academic achievement (as measured by scores on standardized tests) between black and white youth in the United States. The unconditional gap (i.e. the simple difference in average scores between the two groups) is moderate in kindergarten but widens to almost a full standard deviation in later grades.

The substantive question considered by Rothstein and Wozny is whether the black-white test score gap is smaller (or possibly even zero) once we condition on family financial resources. That is, is the difference in average test score between black and white students with the same level of resources smaller than the unconditional difference?

Suppose that each individual in the population (of U.S. grade 3 students) is characterized by the observable variables (s, b, p, c), where s_i is person i's test score, b_i is an indicator for whether she is black, c_i is the current household income of her family, and p is the permanent household income[1] of her family.

We will assume that:

$$
\begin{aligned}
s &= \beta_0 + \beta_1 b + \beta_2 p + u && \text{where } E(u|b,p) = 0 & (1)\\
s &= \gamma_0 + \gamma_1 b + \gamma_2 c + v && \text{where } E(v|b,c) = 0 & (2)\\
p &= \delta_0 + \delta_1 b + \delta_2 c + w && \text{where } E(w|b,c) = 0 & (3)
\end{aligned}
$$

We will also assume that current income is just permanent income plus some random "noise" that is unrelated to any other variable:

$$ c = p + e \qquad \text{where } e \text{ is independent of } (p, b, s). $$

One useful implication of this is that $E(u|b,c) = 0$.

Finally, we add some notation: Let $\theta = (\beta_0, \beta_1, \beta_2, \gamma_0, \gamma_1, \gamma_2, \delta_0, \delta_1)$.

a) The test score gap conditional on current income is defined as:

$$ GAP1(C) = E(s|b=1, c=C) - E(s|b=0, c=C) $$

Find $GAP1(C)$ in terms of the elements of θ

b) The test score gap conditional on permanent income is defined as:

$$ GAP2(P) = E(s|b=1, p=P) - E(s|b=0, p=P) $$

Find $GAP2(P)$ in terms of the elements of θ

c) Previous studies have measured $GAP1(C)$, the gap conditional on current income and found it wasn't much smaller than the unconditional gap. Rothstein and Wozny would like to measure $GAP2(P)$, the gap conditional on permanent income. However, they cannot estimate it directly because there are few data sets with both good test score data and good income data. They have two data sets, one with data on (s, b, c) and another with data on (b, c, p). Both data sets are different random samples from the same population and exhibit nontrivial variation in all variables (i.e, there are no exact linear dependencies). Let $\hat{\gamma} = (\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2)$

---

[1] You may have heard Milton Friedman's concept of "permanent income" described in one of your other classes. Here's the idea: over short time periods people experience substantial income fluctuations due to temporary unemployment, time spent in school, parenting of small children, bonuses, etc. Friedman argued that people smooth their consumption over time, and so the primary determinant of current spending is not current income but rather the level of spending that can be sustained over one's lifetime given one's expected lifetime income path. For example, most university students have current income well below the poverty level, but much higher permanent income.

be the OLS estimator for $\gamma = (\gamma_0, \gamma_1, \gamma_2)$ using the first sample, and let $\hat{\delta} = (\hat{\delta}_0, \hat{\delta}_1, \hat{\delta}_2)$ be the OLS estimator for $\delta = (\delta_0, \delta_1, \delta_2)$ using the second sample. Are these estimators consistent?

d) Using the elements of $\hat{\gamma}$ and $\hat{\delta}$, construct a consistent estimator for $GAP1(C)$.

e) Using the elements of $\hat{\gamma}$ and $\hat{\delta}$, construct a consistent estimator for $GAP2(P)$.

This one is hard, so here's a hint: plug equation (3) into equation (1), and then rearrange so you have something that looks like equation (2).

f) The estimates from the data are:

$$
\begin{aligned}
\hat{\gamma}_1 &= -0.426 \\
\hat{\gamma}_2 &= 0.205 \\
\hat{\delta}_1 &= -0.306 \\
\hat{\delta}_2 &= 0.524
\end{aligned}
$$

Using these estimates and your previous results, construct numerical estimates of $GAP1(C)$ and $GAP2(P)$.

g) Based on these results is the test score gap conditional on permanent income smaller than the gap conditional on current income?