Economics of Education — Frank Levy
14.48J/11.126J — Spring, 2006

Final Exam - Answer All Questions

1) (20 points) Each week, the U.S. Professional Golfers' Association (PGA) publishes an updated ranking of the world's best professional golfers. These are the top seven golfers in the PGA ranking based on professional golf tournaments through May 14, 2006:

1 - Tiger Woods, 2 - Phil Mickelson, 3 - Retief Goosen, 4 - Vijay Singh 5 - Jim Furyk
6 - Ernie Els, 7 - Sergio Garcia,

a) (10 points) Briefly discuss the likely methods the PGA uses to construct their weekly ranking. Explain why you think the PGA ranking of top golfers is more or less accurate than the *U.S. News and World Report* ranking of the top U.S. colleges and universities.

Answer: While each tournament has multiple players, we can think of a tournament as a set of head-to-head competitions and so the PGA ranking likely focuses on who beats who. In addition, the order of finish in different tournaments may get different weights depending on the number of competitors, the tournament's prestige (which may determine how hard people try, etc).

This ranking is likely to be more accurate than the U.S. News Ranking of universities since it is unambiguous that a professional golfer's only goal is winning tournaments. The University, by contrast, has multiple goals, some of which, like the quality of teaching, are not easy to measure and so the U.S. News Rankings are more based on surrogate measures that do not get exactly what you want.

A different way of thinking about it is that we have seen that universities can manipulate the U.S. News rankings by encouraging more applicants, by doing strategic admissions, etc. (none of which are likely to change the university's quality) while if my description of the PGA rankings is correct, it is not clear how those rankings can be manipulated.

b) (10 points) Both the PGA ranking and the *U.S. News* ranking are examples of *commensuration* - a process of applying the same numerical yardstick to different units (different golfers, different universities) in order to rank them. Based on the topics we have discussed this semester, identify a second important education example of commensuration and explain its advantages and limitations.

Answer: The major example is the use of statewide testing programs that are part of the standards movement. The advantage of these tests is that it helps to clarify a school's objective function and creates information that is of some use to parents. The disadvantage is that the tests may measure only part of what you would like to accomplish in education but preparing for the tests is likely to absorb much of a school's efforts ("What you test is what you get.) A secondary disadvantage is that the test scores have a lot of year-to-year fluctuation which can be a problem if too much is riding on them.

2) (20 points) Congress is currently debating U.S. immigration policy including whether the nation should maintain a large guest worker program that would allow significant numbers of Latin American citizens to work here on a temporary basis.

a) (10 points) We know from basic economic theory that when labor becomes scarce and expensive, employers have an incentive to substitute capital for labor. If Congress fails to renew a guest worker program, to what extent will employers be able to directly substitute computers for the work previously done by Latin American guest workers? Most of the work? Some of the work? Not much of the work? Carefully explain your logic and illustrate your answer with some specific examples.

Answer: I would think that not much substitution is possible. A significant fraction of immigrant labor works in lower wage physical and service jobs – restaurants, meat packing plants, domestic work, janitorial work, construction etc. This work is characterized by non-routine physical activity and, in some cases, face-to-face transactions. Both kinds of tasks are hard to express in rules and so hard to computerize.

b) (10 points) During the semester, we argued that computers play a role in determining the earnings gap between college graduates and high school graduates. Does this argument have any relevance for the demand for guest workers? Explain why or why not.

Answer: Ultimately, you would think that there is relevance. When computers substitute for workers in assembly line, clerical and other similar rules based jobs, the displaced workers have to find work elsewhere in the economy – i.e. they either have to move up the wage scale (which requires additional skill) or move down. If the supply of workers moving down gets large enough, some will begin to complete in the same occupations now dominated by immigrants.

(There may be a different answer here about high school drop outs and high school graduates not really competing – you should use your judgment on that and give good points for sensible answers ot this kind).

3) a) (15 points) Construct an argument supporting very early childhood education (ages 6 months to 5 years) as a central piece of U.S. educational policy. Back up the points you make by references to specific readings and ideas discussed in class.

Answer: The argument by citing the work of Heckman that argues three main points:

- IQ measures have a fair amount of predictive power in both labor market measures (educational attainment, adult earnings even holding education constant), and also other measures like drug use, out-of-wedlock pregnancy, etc.

- IQ measures are set relatively early (e.g. by age 4 or 5).

- But evidence from very early childhood programs like Perry Pre School and Abcderian (not spelled right) show that IQ measures can be raised by such interventions.

- Additional points came from class discussion of the black/white test score gap

    - The discussion of the Fryer Levit paper showing that tests of cognitive function at 8 months showed virtually no black/white differences while IQ scores at 4-5 years showed the standard 1sd gap.

    - Limited evidence on interracial adoptions showing that black children raised in adoptive white homes scores significantly higher on IQ tests than black children raised in adoptive black homes).

    - Both of these piece of evidence point to a need for some kind of environmental intervention in a child's very early years.

b) (5 points) Briefly summarize evidence for the counter argument that we don't yet know enough to spend large sums of money on very early childhood education.

Answer:

- Heckman's evidence for IQ gains is really based on only a couple of studies. In addition, Mike Anderson's work shows the evidence is really only there for girls.

- The possibility, raised in class, that we are not really sure what the 8 month cognitive tests are measuring and it could be that if you could ask the 4 year old IQ questions to the 8 month old infants, you would see the same 1sd black white difference.

4) (20 points) Assume that in the year 2001, the Colorado state legislature recognized that student access to public school computers varied dramatically across the state's counties. Urban and suburban counties averaged three students per computer and most of the computers had internet access. Poorer rural counties averaged 21 students per computer and many of the rural school computers lacked internet access.

Members of the legislature believed that lack of computer and internet access was depressing achievement test scores of rural students. Correspondingly, the legislature funded a program that was supposed to purchase computers, software and internet access for "computer poor" counties.

It is now 2006 and the legislature wants to set up an evaluation of the computers' impact. They have turned to you to do the job.

a) (10 points) When administrators establish a program like this, problems often arise that limit the possibility of doing a good evaluation. For example, suppose that you have heard anecdotes that some rural teachers who received the state-funded computers and software used their own funds to purchase in-service training on how to use computers in the classroom. If that actually happened, would it prevent you from evaluating the computers' impact on test scores? Explain why or why not. Describe two other problems (other than data problems) that might prevent you from doing a good evaluation of the computers' impact.

Answer: If some teachers purchase training, it does not invalidate the evaluation. The "treatment" we want to evaluate in this case is putting computers into the schools with the understanding that the state has no ability to control how the computers are used (including whether teachers buy training). In this sense, the evaluation is similar to the Rouse analysis of Milwaukee (the policy was whether the kid won a chance to go to private school, whether or not the used the chance) or the Dynarski evaluation of the Georgia and Arkansas scholarship programs (the policy is making the scholarship available, whether or not students actually used them) or the DuFlo evaluation of the Indonesian Schools (the policy was building schools in particular areas whether or not kids actually attended them.)

If training is effective and only some teachers buy it, results may be statistically insignificant since classrooms where the teachers got training may show positive results and classrooms where teachers didn't get training may show no results (i.e. a high variance). But that would be an accurate representation of reality – that simply putting computers in the classroom does not guarantee results.

Some other factors that might undermine an evaluation: Unknown to you, the "computer rich schools" started additional programs of their own. The computers in the computer-poor schools caused new students (either high achievers or low achievers) to move into the area, etc. One factor that would **not** invalidate the experiment was teachers putting computers in the closet. Again, we are not evaluating "how much can computers raise test scores under ideal conditions" but rather "how much can the state raise test scores by buying computers for schools and dropping them off at the schoolhouse door.".

b) (10 points) Assume that none of these problems actually arise. Explain how you would structure an evaluation of computers' impact on test scores - the data you would need, the way you would use the data, and so on.

Answer: The question is: Who is your control group? One possibility are kids in the urban/suburban Colorado schools. Presumably, they start at a higher level of achievement than the rural kids (which is why the legislature passed the program). So the evaluation here would have to be difs-in-difs: comparing the changes over time in test scores for four groups of kids:

a) Rural kids who went through school before the computers were put in rural schools

b) Rural kids who went through school after computers were put in rural schools.

c) Urban/Suburban kids who went through schools before computers were put in RURAL schools.

d) Urban/Suburban kids who went through schools after computers were put in RURAL schools.

As with the examples we would expect the gap between (b) and (a) to be larger (more improvement) than the gap between (d) and (c)

Alternatively, you could think about trying to find a control group using rural counties in other states that were similar in the test they took, their test scores, etc., but they had no computer program. If we could find such a control group, then we could again do difs-in-difs or put this into a single regression, etc.

5) (20 points) For many years, educational production functions have had a central role in economists' study of education. As in production function analysis for private sector firms, the idea is to collect data from schools to estimate the relationship between a school's inputs and a particular school output – normally some kind of student test score.

In both firms and schools, the estimated relationship between inputs and an output depends on several factors. There is an underlying technology that specifies the maximum amount of a particular output that can be achieved with a particular combination of inputs. There is also the behavior of managers, workers, etc. that determines whether the maximum of a particular output is actually achieved. For example, technology may say that an office worker has the potential to write 45 pieces of correspondence in a day, but this output will be far lower if the worker is told to answer incoming telephone calls as well as write correspondence.

a) (6 points). When economists use private sector data to estimate a production function - for example, the production function for automobile plants - what assumptions are they making about the manager and worker behavior that generated the data? What are the implications of this behavior for the estimated inputs-output relationship?

Answer: They assume that the output of the production function is the firm's focus and that profit maximization is insuring that everyone in the firm is working to get the maximum amount of that output from the available inputs. In other words, what you observe is the maximum that the technology allows.

b) (7 points) Explain how the principal and teacher behavior that generate data for a standard educational production function may differ from the behavior in a private sector firm. What are the implications of these differences for the estimated educational production function?

Answer: The most important factor is that it is not clear schools are trying to maximize achievement test scores. They may be emphasizing skills that the tests do not capture; they may be trying to keep kids from dropping out; they may want to avoid making waves, etc.. Any of these things will cause schools to not reach the potential relationship between inputs and outputs.

c) (7 points) Use your answers in (a) and (b) to explain why it makes sense to run careful experiments on the impact of class size (e.g. Krueger's STAR experiment in Tennessee) or to carefully describe a successful model for in-service training of tenured teachers (e.g. Elmore's study of New York City District 2). In other words, why haven't schools already discovered this knowledge in the course of their normal behavior?

Answer: You need experiments like these to understand the shape of the “potential” educational production function. For example, in the case of the STAR experiment, teachers knew they were part of an experiment and that lots of attention would be paid to the children’s test scores. Given that teachers are focused on the test scores, the experiment can give us some evidence on whether teachers in small classes can do a better job raising test scores than teachers in larger classes. This is particularly true since the experiment ran over multiple years giving teachers a chance to adjust.

In a normal (non-experimental) situation – e.g. the various studies summarized by Hanushek – this relationship may be hard to discover statistically because small classes may be correlated with other (omitted) variables like low achieving students; teachers may not be fully focused on test scores and so there is no reason for them to discover the relationship, etc.

****