University of Illinois | Department of Economics
Fall 2005 | Roger Koenker

Economics 508: Applied Econometrics
Problem Set 4

The written version of this problem set is due Thursday, November 10. There will be the usual formal class presentation of results, stay tuned to the class web page for assigned question numbers.

Background

This problem was inspired by an *AER* paper by Stephan and Levine (1991) on the academic labor market in physics. Unfortunately, their data is proprietary so our data are fictional. Since the papers Divide and Koenker(1991,1994), and Koenker (1995) have been distributed in past years I decided that they should be distributed again this year – this time on the web. There are versions on the class web pages in pdf. You may use these papers as a place to begin to formulate your strategy, *but they should certainly not be regarded as a definitive analysis.* In particular you will note they are based on somewhat different data than that distributed this year.

The academic labor market is unusual in many respects, one of which is that there are wide disparities in (research) productivity over the life cycle and across individuals and considerable concomitant salary dispersion. In this problem we consider detailed data on phuzicists to determine the nature of these life cycle effects. Since phuzics is a rather arcane field, there are only a few research journals, all of which are, essentially, equally irreputable. Therefore we will measure research productivity by the annual number of pages published. To evaluate joint publications we employ Thumb's First Rule that a coauthor of a joint paper of p pages with n authors deserves credit for $p/\sqrt{n}$ pages.

In the sample we have phuzicists of several vintages employed in positions with a considerable range of research expectations. It is also plausible that the vintage of a phuzicist's Phd training may be an important determinant of productivity due to subsequent periods of rapid paradigm-shift. Fortunately, we observe productivity over a rather long period since 1945, and hence we are potentially able to distinguish vintage and life-cycle effects.

The data consists of a panel on individual phuzicists' research productivity, salary, and other characteristics. The raw data consists of observations on individuals; there are 11 variables:

$i \sim$ person identifier
$t_{0i} \sim$ year of Phd - 1900
$t \sim$ current year - 1900
$x_i \sim$gender (==1 if female)
$d_{it} \sim$ indicator of employment in research university (1 if person i is in a research position in year t, and 0 otherwise.)
$y_{it} \sim$ page equivalents in current year
$r_i \sim$ rank of Phd granting University

$Y_{it} \sim$ discounted* cumulative page equivalents
$s_{it} \sim$ current annual salary in \$1,000/yr

A new aspect of the data is the availability of the variable r_i which is the rank of the ith individual's Phd granting University at the time of the Phd. While such rankings are notoriously flawed, it is of obvious interest to investigate the extent to which individuals productivity depends upon the perceived quality of their Phd program. Note also however that since unobserved individual effects on productivity are likely to be also correlated with these ranks, this variable should probably not be treated as exogenous.

What Is to Be Done?

Write a brief research paper (*less* than 5 pages) in which you use the data described above to answer some or all of the following questions.

(1) What is the shape of research productivity in phuzics over the life cycle?
(2) Does "the phuzical revolution" of the 1960's appear to make pre-60's Phd's less productive than their post-60's colleagues?
(3) How much is a page of phuzics research worth as measured by the resulting discounted stream of academic salary? Note that this may depend upon the point in the life-cycle that the page is published, what the accumulated research output is to date, and other factors. Be as explicit as possible here giving examples of "representative individuals" if necessary.
(4) Does the gender of the phuzicist influence productivity or salary and if so, by how much?
(5) The distinguished statistician Emmanuel Parzen has argued that economics is "gradually becoming more scientific" because publication is becoming more concentrated in the hands of fewer researchers. Test this hypothesis for phuzics.
(6) How does the rank of an individuals Phd program influence his/her subsequent research productivity?

Some Hints

Obviously, all six questions are rather open-ended. Your objective should be to narrow them down to an explicit statistical formulation which you can "take to the data." For questions (1) and (2) you might start with a model like

$$logy_{it} = \sum_{s=1}^{q} \rho_i \log y_{it-s} + f(t, t_{0i}, t - t_{0i}, r_i) + u_{it}$$

where t_{0i} is the Phd date of person i. f is some simple function of the current year, Phd vintage, Phd age, and rank of Phd program. Note the natural choice of linear f is fraught with some identification problems. Of course there is also the question of whether u_{it} should include a fixed or random individual effect. To capture the effect of the rank of the Phd program, it might be reasonable to assume that the effect declines over time, and of course lower rank would seem to suggest higher productivity so incorporating a variable like $1/((t - t_{0i})r_i)$ and expecting a positive

*Here we use Thumb's Second Rule that phuzics research depreciates at 8% per year. One might go about verifying this by looking at citation records.

coefficient might be reasonable. Note also that high ability people are likely to attend a highly ranked Phd program.

For questions (3) and (4) a simple initial salary model might look like

$$\log(s_{it}/s_{it-1}) = \gamma \log(Y_{it}/Y_{it-1}) + \text{ other\_effects } + \text{ noise}$$

Caveats

As in any empirical research the objective is to propose and defend an explicit model and associated estimation and inference methods employed to answer the substantive economic questions posed. To phuzicists, the middle two questions are unquestionably *most interesting*. Unfortunately, their answer is complicated by the fact that salary depends on aspects of performance other than measured research and these aspects, one might expect, would be correlated with research performance. While some control over this effect is available though the introduction of individual fixed effects, this creates obvious problems when the effect of time-invariant characteristics like gender are required. Despite these problems, or perhaps because of them, you should recognize that the answer to question (3) deserves more attention than preceding ones. In grading the problem set the four questions will be assigned relative weights 10, 5, 20, 5, 10, 10 respectively, and you should assign your marginal effort to them accordingly.