ECMT 572
Econometrics II

Yoosoon Chang
Spring 2020

Empirical Problems

Due Monday, March 1

1. This exercise makes use of the relationships between log wages, education, gender, married and race. Note that gender, married, and race are dummy (0-1) variables. The earnings data set provided is obtained from Current Population Survey (CPS) conducted by Census Bureau. Use for gender the variable “female” which recodes the gender variable into female/male (1/0), and use for race “black” which recodes the race into black/white (1/0).

(a) Graph the **level** of hourly earnings and **log** of hourly earnings. Describe the shape of the graphs. (we will be graphing empirical density functions).

(b) Regress hourly wage (wage) on education (educ), experience (exp) and its square (expsq), and get the residuals (call it resid1). Also regress log(wage) on educ, exp and expsq, and get the residuals (call it resid2). Then graph resid1 and resid2, and compare them.

(c) What do you conclude from the exercises in (a) and (b)? You can see how well the fitted residuals from a regression using log wages closely approximate a normal distribution, even relative to the distribution of just log wages themselves.

(d) Consider first the means of log wages for whites and blacks. Then construct a test for the difference in these two means being significantly different from zero. A simple way to construct this test is via a regression of log wages on a constant and the race dummy. The coefficient on the race dummy is exactly the difference in the sample average of log wages for blacks and whites. What is the relevant t-test here?

(e) Critics might argue that the race difference in earnings in part (d) might be attributable to racial differences in “skill”, rather than to say, pure discrimination. For our purposes, let us say “skill” can be appropriately measured by the vector of variables, such as education, experience, and experience squared. Test the hypothesis that blacks and whites of the same skill level earn the same wage on average. Is most of the differential in wages explained by differences in observed skill levels? If not what can you conclude?

2. Now consider the gender wage differentials in addition to the racial differences studied in question 1.

(a) Estimate an equation that will allow you to determine whether males and females of the same skill level have different log earnings. What do you conclude? Now augment the model to test the hypothesis that black women are not discriminated against any more than white women. State formally the null hypothesis and conduct the test.

(b) Next consider the following four groups of people: single males, married males, single females, and married females. Controlling for skill, estimate an equation that allows each of the four groups to have a different intercept. Test the hypothesis that there is no difference between the four groups. Also test the hypothesis that there is no difference between single males and single females. What do you conclude?

3. A group of law school students have approached a certain professor about the use of a procedure called the "Urn Model" in a discrimination lawsuit (this procedure, by the way, is actually used in legal studies of discrimination). In such cases, we normally run a regression

$$\log(\text{wage}) = \alpha S + X\beta + \varepsilon \tag{1}$$

where S is a dummy variable indicating sex (say, 1 if the individual is female, and 0 if male) so it is just an $(n \times 1)$ vector, and X is an $(n \times (k-1))$ matrix of additional regressors (including a constant), and α is the (scalar) coefficient on the sex dummy, and β is the $(k-1) \times 1$ coefficient vector on the X's, and look to see if the estimate of α is significantly different from 0. The "Urn Model" instead proposes to just run a regression of log wages on X:

$$\log(\text{wage}) = X\pi + v \tag{2}$$

and then run a regression of the *fitted residuals* from the regression (2) on the sex dummy variable S:

$$\hat{v} = \delta + \gamma S + u \tag{3}$$

where δ is just the coefficient on the intercept, and $\hat{\gamma}$ is the "new" estimate of the gender wage differential from the "Urn Model".

(a) Derive the estimators for α and γ obtained by applying OLS to the regression equations (2) and (3).

(b) Using the results from part (a), write the ratio of $\hat{\alpha}$ to $\hat{\gamma}$ in terms of the R^2 from a regression of S on X. As a result, what can you say <u>definitely</u> about the relative magnitudes of these two estimates of the sex differential?

(c) Using the results you obtained from parts (a) and (b), in addition to what you know of good econometric practice, comment on the merits of the use of the Urn Model as a device to learn something about the "facts". Do you think it is a good idea?

4. It is often claimed that "differencing" the regression model will reduce the degree of autocorrelation. Consider this claim by using as the model:

$$y_t = \beta x_t + u_t \quad \text{and} \quad u_t = \alpha u_{t-1} + \varepsilon_t$$

where (ε_t) is white noise with mean zero and variance σ^2, and $|\alpha| \leq 1$. We want to compare the autocorrelation of (u_t) in this model with the autocorrelation of (v_t) in the following differenced model:

$$y_t - y_{t-1} = \beta(x_t - x_{t-1}) + v_t$$

where $v_t = u_t - u_{t-1}$. Answer the following:

(a) Show that the autocorrelation of u_t, call it $\rho_u(k)$, in the original model is α^k.

(b) Find the autocorrelation of v_t at the first lag, i.e., $\rho_v(1)$, in the differenced model.
(c) Evaluate the claim by comparing $\rho_u(1)$ and $\rho_v(1)$ over the range of α.

5. We are given data for a sex discrimination case in which it is alleged that female professors at a given College of Medicine (COM) received unfairly low salaries given their qualifications. The data consist of the following variables (the first six variables are just department dummies):

Biochem	=1 if in Biochem Dept; =0 otherwise
Physiol	=1 if in Physiology Dept; =0 otherwise
Genetic	=1 if in Genetics Dept; =0 otherwise
Pediatr	=1 if in Pediatrics Dept; =0 otherwise
Medicin	=1 if in Medicine Dept; =0 otherwise
Surgery	=1 if in Surgery Dept; =0 otherwise
Sex	=1 if Male; =0 if Female
Clin	=1 if clinical emphasis; =0 otherwise
Cert	=1 if board certified; =0 otherwise
Prate	Number of publications/years of experience
Exper	Years since obtained MD
Assistn	=1 if an Assistant Professor; =0 otherwise
Associa	=1 if an Associate Professor; =0 otherwise
Fullprof	=1 if a Full Professor; =0 otherwise
Lnsal84	=log(Salary in 1984)

(a) If you were an "Expert Witness" on this case, what economic and statistical justification would you give to the jury for working with **log** of salaries rather than just the salaries themselves? (You can give more than one of each, but to receive full credit, you must give at least one reason that justifies this transformation on economic grounds, and one which is more statistical. In so doing, you may assume that the salary structure for this COM is what we see in the population at large - i.e, use your knowledge from your previous problem set work with the CPS data).

(b) One of the "experts" has run the following regression in which he is trying to decide if men and women receive the same pay versus the idea that women receive lower pay than men. What is your conclusion? Be sure to state precisely what you are testing. Does this appear to be a pretty good model? Interpret what the -0.7110 on the pediatrics department dummy means.

lnsal84	Coeff.	Std. Err	t-ratio
sex	0.0638	0.0333	1.914
biochem	-0.8648	0.0613	-14.113
physiol	-1.0326	0.0610	-16.925
genetic	-0.6971	0.0536	-13.014
pediatr	-0.7110	0.0357	-19.929
medicin	-0.3660	0.0387	-11.855
cert	0.1854	0.0214	8.674
clin	0.1591	0.0410	3.883
prate	-0.0238	0.0173	-1.377
exper	0.0292	0.0053	5.502
expersq	-.000335	.000149	-2.242
assistn	-0.1908	0.0289	-6.603
associa	-0.0811	0.0237	-3.427
constant	12.0696	0.0749	161.130

Number of Observations = 261, $RSS = 4.3528$, $R^2 = 0.9352$

(c) In terms of log(salary84), how much does the board certified female assistant professor in the surgery department with 5 years of experience and clinical emphasis get paid on the average? How about the male professor with exactly the same qualifications? What is the difference in the average salaries in actual dollars of the male and female professors in this case?

(d) The "expert" then wants to see if **rank** in the regression in part (b) is significant. In doing so, he runs the following regression and sees that the increase in the R^2 is only about 0.012 when the rank indicators are added to the model. Is this a significant change? (see the output on the next page.)

lnsal84	Coeff.	Std. Err	t-ratio
sex	0.1083	0.0352	3.079
biochem	-0.8144	0.0657	-12.397
physiol	-0.9617	0.0649	-14.821
genetic	-0.6560	0.0575	-11.413
pediatr	-0.6798	0.0381	-17.847
medicin	-0.3401	0.0331	-10.291
cert	0.1733	0.0228	7.617
clin	0.1401	0.0441	3.175
prate	-0.0347	0.0186	-1.870
exper	0.0495	0.0046	10.741
expersq	-.000759	.000144	-5.259
constant	11.8245	0.0700	168.842

Number of Observations = 261, $RSS = 5.1224$, $R^2 = 0.9237$

(e) Our "expert", possessing now notices that the coefficient on the sex indicator now has a much higher t-statistic. First of all, what economic interpretation would you give to the jury in terms of what this coefficient *means*. Secondly, what reasons do you provide the jury with *why* the t-statistic increased so much relative to the model in part (b)? Thirdly, are the data consistent with a null hypothesis that the coefficient on sex is 0.2083?