ECO671, Spring 2004, Final Exam
Prof. Bill Even

1. (30 points) Below are the results of a random effects and a fixed effects regression that estimates the rate of return earned by a pension plan as a function of the number of participants, a dummy indicating whether the plan is participant directed, and the share of assets invested in three categories (omitted group is "other"), and year dummies. The data are from the years 1990 through 1998. Both the random and fixed effects model group observations according to the specific pension plan (i.e. einpn identifies a specific pension plan over time.)

Random-effects GLS regression
Number of obs = 18418
Group variable (i): einpn
Number of groups = 4812
R-sq: within = 0.1250 Obs per group: min = 1
between = 0.1061 avg = 3.8
overall = 0.1205 max = 9
Random effects u_i ~ Gaussian
corr(u_i, X) = 0 (assumed)
Wald chi2(13) = 2535.61
Prob > chi2 = 0.0000

Fixed-effects (within) regression
Number of obs = 18418
Group variable (i): einpn
Number of groups = 4812
R-sq: within = 0.1276 Obs per group: min = 1
between = 0.0797 avg = 3.8
overall = 0.1008 max = 9
corr(u_i, Xb) = -0.2446
F(13,13593) = 153.00
Prob > F = 0.0000
F test that all u_i=0:
F(4811, 13593) = 1.19
Prob > F = 0.0000

| | Random effects | | Fixed effects | |
|---|---|---|---|---|
| | Coefficient | t-statistic | Coefficient | t-statistic |
| # of participants | 1.12E-07 | 0.93 | 2.39E-07 | 0.63 |
| participant directed | 0.008 | 1.99 | 0.01 | 1.74 |
| % of assets in: | | | | |
| government bonds | -0.051 | -2.32 | -0.08 | -2.13 |
| stocks (other than employer) | 0.078 | 8.72 | 0.15 | 7.32 |
| employer stock | 0.099 | 15.58 | 0.22 | 10.39 |
| Year dummies: | | | | |
| 1991 | 0.194 | 28.41 | 0.19 | 25.86 |
| 1992 | 0.11 | 16.18 | 0.10 | 13.52 |
| 1993 | 0.096 | 14.27 | 0.08 | 10.93 |
| 1994 | 0.003 | 0.48 | 0.00 | -1.11 |
| 1995 | 0.198 | 29.17 | 0.18 | 23.26 |
| 1996 | 0.138 | 20.08 | 0.12 | 15.05 |
| 1997 | 0.205 | 29.05 | 0.19 | 21.85 |
| 1998 | 0.104 | 14.39 | 0.08 | 9.14 |
| _Intercept | -0.01 | -2.53 | -0.04 | -4.58 |
| | | | | |
| sigma_u | 0.08 | | 0.17 | |
| sigma_e | 0.21 | | 0.21 | |
| rho | 0.13 | | 0.39 | |

a. Based on the information provided, is the random effects model an appropriate specification? Why or why not? Describe the basis for your answer and the specific hypothesis that is tested to determine whether it is "appropriate".

b. Based on the information provided, is the fixed effects model preferred or the random effects model? Why or why not? Describe the basis for your answer.

c. In the fixed effects model, define $a_i$ as the pension-specific intercept for pension i. What evidence is there that $a_i$ is correlated with the percent of assets that pension i invests in company stock? How can you deduce the sign of this correlation from the results that are presented?

d. Examining the fixed effects results, notice the f-statistic from the fixed effects regression.

F test that all u_i=0:
F(4811, 13593) = 1.19
Prob > F = 0.0000

Explain how this test statistic was constructed and why it has (4811, 13593) degrees of freedom.

ECO671, Spring 2004, Final Exam
Prof. Bill Even

2. (15 points) In a study of immigrant earnings, Barry Chiswick used **cross-sectional** data to estimate an earnings function for natives (U.S. born people) and immigrants. In this study, he found that the effect of age on earnings was larger for immigrants. He argued that this implied that, while immigrants may start at lower salaries and may be less productive than natives, they eventually become more productive as they acquire language and other skills. George Borjas followed up on this study and used **panel data**. For simplicity, assume that he estimated a fixed effects model. He found immigrants did not illustrate higher wage growth than natives.

Assuming it is not simply "different data", what can explain the differences between the conclusions of these two studies? Provide an econometric basis for your explanation and an intuitive explanation of the underlying relationships in the data that could cause the difference in the estimated effects.

3. (30 pts) Suppose that you work for a telephone company and have been asked to examine the effects of a new calling plan. Without the plan, a customer pays $.10 per minute of long distance but has no fixed monthly fee. With the plan, the customer pays $.05 per minute but pays $5 per month. Suppose that you do an OLS regression as follows:

$$minutes_i = a_0 + a_1 * CALLPLAN_i + a_2 * income_i + e_i$$

where minutes is the number of minutes of long distance per month, CALLPLAN is a dummy that equals one if the customer buys the second plan (i.e. $5 per month plus $.05 per minute), and income is the customer's income. The subscript i identifies customers.

a. Explain why the above regression would likely yield a biased estimate of $a_1$. What is the likely direction of the bias and why?
b. Suppose that higher income people are more likely to use lots of minutes. How and why would this cause a bias in the estimated coefficient on income? Explain.
c. Explain how you could correctly estimate the effect of the calling plan on minutes used. Provide a DETAILED explanation of the process and define your variables. Assume that you have only cross-sectional data.
d. Suppose the company is interested in forcing all their customers on to the calling plan. Explain precisely how you would estimate
   i. the change in minutes used for people who are not currently on the calling plan.
   ii. total minutes for people who do not previously use the calling plan when they are forced onto the calling plan.

4. (25 points) Suppose you have been hired by a bank to examine default risk of loan applicants. You have been provided with the loan records of thousands of previous applicants and have information on the percentage of the loan that was paid off (the "payoff rate"). You also have information on several attributes at the time the person applied for the loan, such as credit rating, marital status, and income. The loan officers who decided whether to extend these loans had the same information available, as well as some information that was not kept in the data base. The loan officers did reject some loan applications and you do not observe pay-off information for applicants who were denied credit.

a. Suppose you regress the percent of loan paid off on the loan appliants characteristics. You then use this regression to estimate the percent of loan that a new pool of applicants will pay off. Is your estimate for the new pool of applicants likely to be too high or too low? Why?

b. How could you improve your estimate of the "pay-off" rate for a pool of applicants who
   i. have been approved for a loan.
   ii. have been reject for a loan.
Explain with specific details on the methods you would employ.