Economics 361
Problem Set #6

Jun Ishii *
Department of Economics
Amherst College

Fall 2022

Question 1: Minimax Play at Wimbledon

This question requires you to read "Minimax Play at Wimbledon," by Mark Walker and John Wooders, published in the December 2001 volume of the *American Economic Review*, and the associated lecture handout "Hypothesis Testing and Tennis."

(a) Briefly explain why the authors believe that their test, using data from professional tennis matches, is a "better" test of the minimax theorem than earlier tests using data from minimax experiments (experimental data)?

ANS: Read the following paragraphs from pp.1521-1522

> We begin from the observation that games are not easy to play, or at least to play well. This is especially true of games requiring unpredictable play. Consider poker – say, five-card draw poker. The rules are so simple that they can be learned in a few minutes' time. Nevertheless, a player who knows the rules and the mechanics of the game but has little experience actually playing poker will not play well. Similarly, in experiments on minimax play the rules of the game have typically been simple, indeed transparently easy to understand. But subjects who have no experience actually playing the game are not likely to understand the game's strategic subtleties – they are likely to understand how to play the game, but not how to play the game well. Indeed, it may simply not be possible in the limited time frame of an experiment to become very skilled at playing a game that requires one to be unpredictable.
>
> Professional sports, on the other hand, provide us with strategic competition in which the participants have devoted their lives to becoming experts at their games, and in which they are often very highly motivated as well. Moreover, situations that call for unpredictable play are nearly ubiquitous in sports: The pitcher who 'tips' his pitches is usually hit hard, and batters who are known to 'sit on' one pitch usually don't last long. Tennis players must mix their serves to the receiver's forehand and backhand sides; if the receiver knew where the serve was coming, his returns would be far more effective. Point guards who can only go to their right don?t make it in the NBA. Thus, while the players' recognition of the 'correct' way to mix in these situations

*Office: Converse Hall 315 Phone: (413) 542-2901 E-mail: jishii@amherst.edu

> may be only subconscious, any significant deviation from the correct mixture will generally be pounced upon by a sophisticated opponent.
>
> As empirical tests of the minimax hypothesis, however, sports are generally inferior to experiments. In the classic confrontation between pitcher and batter, for example, there are many actions available (fastball, curve, change-up, in- side, outside, etc.), and the possible outcomes are even more numerous (strike, ball, single, home run, fly ball, double-play grounder, etc.). Difficulties clearly arise in modeling such situations theoretically, in observing players' actions, and in obtaining sufficient data to conduct informative statistical tests.
>
> Tennis, however, provides a workable empirical example: although speed and spin on the serve are important choices, virtually every (first) serve is delivered as far toward one side or the other of the service court as the server feels is prudent, and the serve is for many players an extremely important factor in deter- mining the winner of the point. Moreover, theoretical modeling is tractable (each point has only two possible outcomes: either the server wins the point, or the receiver does); the server's actions are observable (it is easy to see whether he has served to the receiver's forehand or backhand); and data is relatively plentiful (long matches contain several hundred points played by the same two players).

(b) The intuition underlying the Pearson statistic is that one should reject the null hypothesis when the difference between the observed frequency of some event (O_j) and the expected frequency under the null hypothesis (E_j) is large. In other words, reject when ($O_j - E_j$) is too positive or too negative. This suggests a "two-sided" hypothesis test. Yet, hypothesis tests involving the Pearson statistic are almost always one-sided: reject if the test statistic is too large (positive). Explain why.

ANS: Note that the Pearson test statistic involves the sum of *squared* differences between observed and expected frequency of possible events. Therefore, positive differences of a certain magnitude are treated symmetrically with negative differences of a certain magnitude. Therefore, even though the alternative hypothesis allows for both positive and negative differences, what matters in terms of the Pearson test is not the sign of the difference but the magnitude.

(c) See Figure 4 on p.1532. What is the minimum value of the function being drawn? What does that minimum value indicate? Also, for what value of the "receiver's mixture probability on the left" does the function achieve its minimum? What is the relationship between that (argmin) mixture probability and the underlying hypothesis test?

ANS: The drawn function is the power function associated with the authors' hypothesis test at the 5% significance level – i.e. the probability that the chosen test statistic and critical region will result in a rejection of the null as a function of the "true" parameter values. The minimum occurs at the Receiver's Probability on Left of $\frac{2}{3}$, the value under the Null Hypothesis. So the minimum value is 0.05 (or 5%) – the significance level, the fixed probability of a Type I Error.

(d) Use the data in Table 1 on p.1526 to test the null hypothesis of whether Sampras in the 1995 U.S. Open when serving in the "Ad" court utilized a proper mixed strategy. Use the Pearson test statistic and a significance level of 5%. Clearly show your work/steps. Start by showing why the Pearson test statistic value for this test is 1.524 (as stated in Table 1).

ANS: From "Hypothesis Testing and Tennis" handout, the relevant (Pearson) test statistic is

$$\hat{\text{TS}}\text{'} \;=\; \left(\frac{O_1^L - N^L\hat{p}}{\sqrt{N^L\hat{p}(1-\hat{p})}}\right)^2 + \left(\frac{O_1^R - N^R\hat{p}}{\sqrt{N^R\hat{p}(1-\hat{p})}}\right)^2 \;\overset{a}{\sim}\; \chi_1^2 \qquad \text{under } H_0$$

$$\text{where } \hat{p} \;=\; \frac{O_1^L + O_1^R}{N^L + N^R}$$

From Table 1: $O_1^L = 12$, $O_1^R = 28$, $N^L = 20$, $N^R = 37$

By applying numbers above to TS', we can show that $TS' \approx 1.524$

For a significance level of 5%, we reject if $TS' > 3.841$. Therefore, we fail to reject the null.

Question 2: Some Goldberger Problems

Problems **(a) - (c)** are adapted from Goldberger Problem 16.2. You are given a random sample (Y, X) that satisfies the Gauss Markov assumptions. In addition, you are told that $\sigma^2 = 1$ and

$$X'X \;=\; \begin{pmatrix} 4 & -2 \\ -2 & 3 \end{pmatrix}$$

Note that X is a (N x 2) matrix. Let β_1 denote the coefficient before the first variable and β_2 the second variable in X.

(a) Calculate Var($b_1^{ols}|X$), Var($b_2^{ols}|X$), and Cov($b_1^{ols}, b_2^{ols}|X$)

ANS: First, calculate $(X'X)^{-1}$

$$(X'X)^{-1} \;=\; \frac{1}{3*4-(-2)*(-2)}\begin{pmatrix} 3 & 2 \\ 2 & 4 \end{pmatrix} \;=\; \begin{pmatrix} \frac{3}{8} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

As $\sigma^2 = 1$, Var($b_1^{ols}|X$) $= \frac{3}{8}$, Var($b_2^{ols}|X$) $= \frac{1}{2}$, and Cov($b_1^{ols}, b_2^{ols}|X$) $= \frac{1}{4}$

(b) Calculate Var($b_1^{ols} + b_2^{ols}|X$) and Var($b_1^{ols} - b_2^{ols}|X$)

ANS:

$$\begin{aligned} \text{Var}[b_1^{ols} + b_2^{ols}|X] &= \frac{3}{8} + \frac{1}{2} + 2 \times \frac{1}{4} \;=\; \frac{11}{8} \\ \text{Var}[b_1^{ols} - b_2^{ols}|X] &= \frac{3}{8} + \frac{1}{2} - 2 \times \frac{1}{4} \;=\; \frac{3}{8} \end{aligned}$$

Let $t_1 = b_1^{ols} + b_2^{ols}$ and $t_2 = b_1^{ols} - b_2^{ols}$. Let $\theta_1 = \beta_1 + \beta_2$ and $\theta_2 = \beta_1 - \beta_2$

(c) Now answer Problem 16.2 in Goldberger:

> You are offered the choice of two jobs: estimate $\beta_1 + \beta_2$ or estimate $\beta_1 - \beta_2$. If you choose the former, you will be paid $10 - (t_1 - \theta_1)^2$. If you choose the latter, you will be paid $10 - (t_2 - \theta_2)^2$. To maximize your expected pay, which job should you take? What pay will you expect to receive?

ANS: Note that $E[t_1|X] = \theta_1$ and $E[t_2|X] = \theta_2$. So $E[(t_i - \theta_i)^2|X] = \text{Var}[t_i|X]$. Therefore, choosing the job with the higher expected pay is the same as choosing the job with the lower variance: so choose to estimate $\theta_2 = \beta_1 - \beta_2$

(d) This is a slightly re-worded version of Goldberger Problem 20.1.

You are given a random sample (Y, X) that satisfies the classical normal regression model (Gauss Markov + Multivariate Normality) assumptions. In addition, you are told that $\sigma^2 = 2$ and

$$X'X = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$$

There are 32 observations in the sample. The OLS estimates are $b_1 = 3, b_2 = 2$

Test at the 5% significance level the following *joint* null hypothesis

- $H_0 : \beta_1 = 3$ and $\beta_2 = 3$

What is the alternative hypothesis against which you are testing?

ANS:

$$R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I \qquad r = \begin{pmatrix} 3 \\ 3 \end{pmatrix} \qquad b = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

$$(Rb - r) = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \qquad [R\,\sigma^2(X'X)^{-1}\,R']^{-1} = [\,\sigma^2(X'X)^{-1}\,]^{-1} = \frac{1}{2}(X'X) = \begin{pmatrix} \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$$

$$\begin{aligned} \chi^2 \text{ Test Statistic} &= (Rb - r)'\,[R\,\sigma^2(X'X)^{-1}\,R']^{-1}\,(Rb - r) = 1 \\ \text{Critical Value for } \chi^2_2 \text{ for 5\% significance} &= 5.99 \text{ (See Goldberger Appendix Table A.2)} \end{aligned}$$

So we **fail to reject** the null hypothesis against the alternative hypothesis

- $H_a : \beta_1 \neq 3$ **or** $\beta_1 \neq 3$

With $s^2 = 2$ (in lieu of $\sigma^2 = 2$), the test statistic above (where we sub s^2 for σ^2) no longer has a sampling distribution of χ^2_2 but rather $F_{2,30}$. Note: $N - k = 32 - 2 = 30$ The critical value should come from the distribution table for $F_{2,30}$.

(e) This is a slightly reworded version of Goldberger Problem 28.3.

$$\text{Suppose that:}\quad \begin{aligned} y_1 &= \theta + \epsilon_1 \\ y_2 &= 2\theta + \epsilon_2 \\ y_3 &= 3\theta + \epsilon_3 \end{aligned}$$

You observe $\{y_1, y_2, y_3\}$ but not $\{\epsilon_1, \epsilon_2, \epsilon_3, \theta\}$. $(\epsilon_1, \epsilon_2, \epsilon_3)$ are distributed **independently** of each other with the same mean of zero but different variances: $\sigma_1^2 = 4, \sigma_2^2 = 6, \sigma_3^2 = 8$.

Find the minimum variance linear unbiased estimator (MVLUE) of θ. Linear refers to linear function of (y_1, y_2, y_3).

ANS: Consider the following size 3 sample

$$\{(y_1, 1), (y_2, 2), (y_3, 3)\}$$

i.e. $x_1 = 1, x_2 = 2, x_3 = 3$

Note that $E[y_i|X] = \theta x_i$ and $\text{Var}[y_i|X] = \begin{cases} 4 & \text{if } i = 1 \\ 6 & \text{if } i = 2 \\ 8 & \text{if } i = 3 \end{cases}$ and $\text{Cov}[y_i, y_j|X] = 0$ for $i \neq j$ as the ϵ's are distributed independently of each other (and thus the y's after conditioning on X)

So, we essentially have a size 3 sample that satisfies the linearity condition (and full rank) but not spherical errors. We can apply GLS to the sample to obtain MLVUE of θ, by Aitken's Theorem.
Note: $\Omega = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 8 \end{pmatrix}$ and $\hat{\theta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$

Question 3: A Fitness Test for Heteroskedasticity

Consider the following "regression equation" for some observation i of a given sample (Y, X)

$$Y_i \;=\; X_i'\beta + \epsilon_i \qquad \text{note: } X_i \text{ and } \beta \text{ are both } (k \times 1) \text{ vectors}$$

We can express the regression equation in matrix form as

$$Y \;=\; X\beta + \epsilon$$

Let b^{ols} be the OLS estimator of β. X achieves full (column) rank and includes the customary column of ones ("1").

Denote the value of Y predicted by OLS as follows:

$$\hat{Y} \;=\; \begin{pmatrix} \hat{Y}_1 \\ \vdots \\ \hat{Y}_N \end{pmatrix} \;=\; \begin{pmatrix} X_1'b^{ols} \\ \vdots \\ X_N'b^{ols} \end{pmatrix} \;=\; Xb^{ols}$$

Denote the associated residuals as follows:

$$e \;=\; \begin{pmatrix} e_1 \\ \vdots \\ e_N \end{pmatrix} \;=\; \begin{pmatrix} Y_1 - \hat{Y}_1 \\ \vdots \\ Y_N - \hat{Y}_N \end{pmatrix} \;=\; Y - \hat{Y}$$

(a) Explicitly show that $\hat{Y}'e = 0$

ANS:

$$\begin{aligned} \hat{Y}'e \;&=\; (Xb^{ols})'e \;=\; b^{ols\prime}X'e \;=\; \underbrace{(Y'X(X'X)^{-1})}_{b^{ols\prime}}\; X' \;\underbrace{(Y - X(X'X)^{-1}X'Y)}_{e} \\ &=\; Y'X(X'X)^{-1}X'Y - Y'X(X'X)^{-1}X'X(X'X)^{-1}X'Y \\ &=\; Y'X(X'X)^{-1}X'Y - Y'X(X'X)^{-1}X'Y \;=\; 0 \end{aligned}$$

(b) Explicitly show that $Y'Y = \hat{Y}'\hat{Y} + e'e$

ANS:

$$Y'Y \;=\; (\hat{Y} + e)'(\hat{Y} + e) \;=\; \hat{Y}'\hat{Y} + \hat{Y}'e + e'\hat{Y} + e'e \;=\; \hat{Y}'\hat{Y} + e'e$$

Recall from **(a)** that $\hat{Y}'e = 0$ and therefore $e'\hat{Y} = 0$

(c) The result shown in **(b)** is often expressed in words as

> "The total sum of squares (TSS) is equal to the explained sum of squares (ESS) plus the residual sum of squares (RSS)."

Briefly explain why.

ANS: Note that

$$Y'Y \;=\; \sum_i Y_i^2 \qquad \hat{Y}'\hat{Y} \;=\; \sum_i \hat{Y}_i^2 \qquad e'e \;=\; \sum_i e_i^2$$

Let $\bar{Y} = \frac{1}{N}\sum_i Y_i$ and $\bar{\hat{Y}} = \frac{1}{N}\sum_i \hat{Y}_i$ and $\bar{e} = \frac{1}{N}\sum_i e_i$

(d) Explicitly show that $\bar{Y} = \bar{\hat{Y}}$. **Hint:** $\bar{e} \;=\; ?$

ANS:

$$\bar{Y} \;=\; \frac{1}{N}\sum_i Y_i \;=\; \frac{1}{N}\sum_i \left(\hat{Y}_i + e_i\right) \;=\; \bar{\hat{Y}} + \bar{\hat{e}} \;=\; \bar{\hat{Y}}$$

Recall from PS #4 that $\bar{e} = 0$

(e) Explicitly show that the sample variance of Y is equal to the sample variance of $\hat{Y}$ plus the sample variance of e: $\frac{1}{N}\sum_i(Y_i-\bar{Y})^2 = \frac{1}{N}\sum_i(\hat{Y}_i-\bar{\hat{Y}})^2 + \frac{1}{N}\sum_i(e_i-\bar{e})^2$

ANS: Combine **(b)** **(c)** **(d)** with

$$\begin{aligned}
\frac{1}{N}\sum_i(Y_i-\bar{Y})^2 &= \frac{1}{N}\sum_i Y_i^2-(\bar{Y})^2 \\
\frac{1}{N}\sum_i(\hat{Y}_i-\bar{\hat{Y}})^2 &= \frac{1}{N}\sum_i \hat{Y}_i^2-(\bar{\hat{Y}})^2 \\
\frac{1}{N}\sum_i(e_i-\bar{e})^2 &= \frac{1}{N}\sum_i e_i^2-(\bar{e})^2
\end{aligned}$$

A common (but not necessarily compelling) practice for researchers using OLS is to report a "goodness of fit" statistic known as R^2, defined as follows:

$$R^2 \equiv \frac{\sum_i(\hat{Y}_i-\bar{Y})^2}{\sum_i(Y_i-\bar{Y})^2}$$

Heuristically, R^2 reflects how much of the variation in Y is explained by the variation in the (OLS) predicted values of Y. Recall our earlier motivation of OLS as an estimator of the $BLP_{MSE}(Y|X)$ (and the $BP_{MSE}(Y|X)$ with the Linearity Condition).

(f) Explain why $R^2 \in [0,1]$. Explain why R^2 may not be bounded between 0 and 1 if OLS is estimated *without* a constant (i.e. exclude column of ones from X).

ANS:

$$\begin{aligned}
\frac{1}{N}\sum_i(Y_i-\bar{Y})^2 &= \frac{1}{N}\sum_i(\hat{Y}_i-\bar{\hat{Y}})^2+\frac{1}{N}\sum_i(e_i-\bar{e})^2 \\
1 &= \underbrace{\frac{\sum_i(\hat{Y}_i-\bar{\hat{Y}})^2}{\sum_i(Y_i-\bar{Y})^2}}_{R^2}+\frac{\sum_i(e_i-\bar{e})^2}{\sum_i(Y_i-\bar{Y})^2} \\
R^2 &= 1-\frac{\sum_i(e_i-\bar{e})^2}{\sum_i(Y_i-\bar{Y})^2} = 1-\frac{\sum_i e_i^2}{\sum_i(Y_i-\bar{Y})^2} = 1-\frac{\sum_i(Y_i-\hat{Y}_i)^2}{\sum_i(Y_i-\bar{Y})^2}
\end{aligned}$$

Note that b^{ols} is chosen to minimize $\sum_i e_i^2$. Therefore $\sum_i(Y_i-\hat{Y}_i)^2 \le \sum_i(Y_i-\bar{Y})^2$ as b^{ols} can always be chosen to make $\hat{Y}_i=\bar{Y}$. Thus, $\frac{\sum_i(Y_i-\hat{Y}_i)^2}{\sum_i(Y_i-\bar{Y})^2} \le 1$. Note that sum of squares must be non-negative. Thus, $\frac{\sum_i(Y_i-\hat{Y}_i)^2}{\sum_i(Y_i-\bar{Y})^2} \ge 0$. Therefore $R^2 \in [0,1]$

The above result used $\bar{e}=0$. Excluding the column of ones can lead to $\bar{e}\neq 0$. See the first order conditions associated with OLS.

(g) What must the values of $\{Y_i\}_{i=1}^N$ be in order for $R^2=1$? What must the values of $\{\hat{Y}_i\}_{i=1}^N$ be in order for $R^2=0$? For both questions, explain why.

ANS:

- $R^2 = 1$: need $\frac{\sum_i (Y_i - \hat{Y}_i)^2}{\sum_i (Y_i - \bar{Y})^2} = 0$. This occurs only if $Y_i = \hat{Y}_i = X_i' b^{ols} \;\; \forall i$

- $R^2 = 0$: need $\frac{\sum_i (Y_i - \hat{Y}_i)^2}{\sum_i (Y_i - \bar{Y})^2} = 1$. This occurs only if $\hat{Y}_i = \bar{Y} \;\; \forall i$

Halbert White, in a classic 1980 article in *Econometrica*, proposed a "simple" test to see whether a given sample violated the Homoskedasticity condition (same conditional variance across observations) – i.e. test of $H_o : \sigma_i^2 = \sigma_j^2$ for all (i, j) in sample vs. $H_a : \sigma_i^2 \neq \sigma_j^2$ for some (i, j) in sample.[1] The other two Gauss-Markov conditions are assumed to be satisfied.

The steps for calculating the White test statistic for testing heterpskedasticity are as follows:

1. Run OLS on (Y, X) as usual, save the residuals $\{e_i\}_{i=1}^N$
2. Apply OLS on the following regression equation $(e_i)^2 = Z_i'\gamma + \eta_i$ where $\eta_i \equiv (e_i)^2 - Z_i'\gamma$

   where Z_i includes all elements in X_i, their squares, and their cross-products

   (excluding redundancies)
3. Calculate R^2 from the above "auxiliary" regression and multiply it by N to get the test statistic: $TS = NR^2$

White is able to show that, under the null hypothesis of homoskedasticity, the above test statistic has an *asymptotic* distribution of χ^2 with $q - 1$ degrees of freedom where q is the number of variables/elements in Z_i: $NR^2 \overset{a}{\sim} \chi^2_{q-1}$

(h) The derivation of this test statistic and its sampling distribution (when H_o is true) is beyond the scope of this course (but not by too much). However, the intuition for this test statistic is not, especially given the discussion earlier in this question. Provide the intuition.

ANS:

- Recall that $\sigma_i^2 = V(Y_i|X) = V(\epsilon_i|X) = E[\epsilon_i^2|X]$

  (the last part as $E[\epsilon_i|X] = 0$ with Linearity Condition)
- Therefore, we can consider e_i^2 as something of a sample analog for σ_i^2
- If the errors are homoskedastic, we would not expect e_i^2 to vary systematically across the sample as $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_N^2$
- More specifically, we would not expect the variation in e_i^2 across the sample to be explained well by the variation in X_i and other variables that vary across $i = 1 \ldots N$ leading to a low R^2 value for the auxiliary regression (based on earlier discussion of R^2)
- Note that we reject H_0 for high values of R^2 and fail to reject for low values of R^2

Aside: There are many (including your Econ 361 professor and Arthur Golberger) who believe that R^2 is *much* over-valued by the profession. Note that even if $b^{ols} = \beta$ (perfect estimation), R^2 will generally be less than 1. However, R^2 is useful in deriving/calculating some test statistics.

[1] "A Heteroscedastic-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity," *Econometrica*, 1980, pp.817-838

Question 4: Unobserved Ability and Fixed Effects

(a) Briefly explain why this sample is **not** random

ANS: Even if $ln(\text{Wage}_{it}) \mid \text{EXP, ABIL}$ is distributed independently and identically (i.i.d.), $ln(\text{Wage}_{it}) \mid \text{EXP}$ most likely will not be as the relationship between $ln(\text{Wage})$ and EXP will differ across i due to the omitted **time invariant** $ABIL$ – so not identical. Additionally, $ln(\text{Wage}_{it}) \mid EXP$ will be correlated across t within the same i due to the omitted $ABIL$ – so not independent.

(b) Show or explain why

$$E[ln(\text{Wage}_{it}) \mid \text{EXP}] \quad = \quad \beta_0 + \beta_1 \text{EXP}_{it} + \beta_2 E[\text{ABIL}_i \mid \text{EXP}]$$

ANS: You should know the answer by now: application of Law of Iterated Expectations.

(c) Suppose you believe that EXP_{it} and ABIL_i are positively correlated with each other; workers with greater innate ability tend to have more work experience. (e.g. Workers with higher innate ability tend to get work earlier and keep their jobs longer) What does this belief suggest about $\frac{d}{dEXP}\, E[\text{ABIL}_i \mid \text{EXP}]$?

ANS: We would expect $E[\text{ABIL}_i \mid \text{EXP}]$ to rise with EXP. Workers with high experience tend to be those with high ability. So the benefits of high experience tends to be confounded with the benefit of high ability, suggesting $\frac{d}{dEXP}\, E[\text{ABIL}_i \mid \text{EXP}] > 0$. Note: this assumes that both β_1 and β_2 are positive.

Download the data set `wages.csv` from the course website. It contains experience, ability, and wage data for 4 workers over 10 years (sample size of 40). The dataset should be read using the following command: `infile i t exp abil wage using wages.csv`

If you stored `wages.csv` is a directory other than the STATA data directory, put the directory address in front of `wages.csv`. e.g. if the file is in `C:\projects\data` then use:
`infile i t exp abil wage using C:\projects\data\wages.csv`

The variables are

- i indicates the worker (1 through 4)
- t indicates the year (1 through 10)
- EXP indicates the experience for worker i in year t
- ABIL indicates the innate ability of worker i
- WAGE indicates the wage for worker i in year t

(d) Regress Wage on (EXP, ABIL, constant). Report the estimated coefficients and standard errors

	Coef.	Std. Error
exp	0.4685459	0.1704306
abil	0.4520185	0.2769119
constant	5.767644	1.979247

(e) Regress Wage on (EXP, constant) – omitting ABIL. Report the estimated coefficients and standard errors

	Coef.	Std. Error
exp	0.5371249	0.2141389
constant	7.752199	1.821606

(f) Use the STATA Command correl EXP ABIL to calculate the sample correlation between EXP and ABIL. How does this sample correlation help explain the difference between estimated coefficient before **EXP** obtained in **(d)** and in **(e)**?

ANS: The sample correlation between EXP and ABIL is 0.2465. The positive correlation suggests that $E[ABIL|EXP]$ rises with EXP. Therefore, the estimated coefficient before EXP in the shortened regression **(e)** will be biased upward, as the positive relationship between wage and experience will be confounded with the positive relationship between wage and ability.

(g) Use the following STATA Commands to create fixed effects for each worker (d1 - d4)

- gen d1 = 0
- replace d1 = 1 if i == 1
- gen d2 = 0
- replace d2 = 1 if i == 2
- gen d3 = 0
- replace d3 = 1 if i == 3
- gen d1 = 0
- replace d4 = 1 if i == 4

(h) Try to regress Wage on EXP, ABIL, d1, d2, d3, d4, constant. Explain why this fails. (**HINT:** Full rank)

ANS: Note that $d1_i + d2_i + d3_i + d4_i = 1$ as each of the observations in the sample correspond to one and only one of the four workers. So the 4 dummy variables are perfectly multicolinear with the intercept. The full rank condition is violated and "$(X'X)^{-1}$" does not exist.

(i) Regress Wage on EXP, d1, d2, d3, d4 (no constant). Compare these estimates to those of **(d)** and **(e)**. Do fixed effects seem to help address "omitted variables bias" ?

	Coef.	Std. Error
d1	9.68758	2.020644
d2	7.311607	1.719493
d3	8.103567	1.867271
d4	8.867807	2.17843
exp	0.4548537	0.1820104

The inclusion of the fixed effects do lower the coefficient estimate for EXP, closer to the estimated value for the long regression that includes ABIL. In expectation, the inclusion of the fixed effects should lead to the coefficient estimate for EXP to have the same expected value in this regression as in the long regression. But in small samples, there can be differences.

(j) Take a stab at the intuition for the result in **(i)** concerning fixed effects. **Hint:** See **(b)**

ANS: The estimated fixed effects are essentially estimates of $\beta_0 + \beta_2 E[\text{ABIL}_i \mid \text{EXP}]$ for each worker. i.e. The estimated coefficient before $d1$ is an estimate of $\beta_0 + \beta_2 E[\text{ABIL}_i \mid \text{EXP}]$ for worker 1. As the omitted variable, ABIL_i, does not vary across years for a worker (and $E[\text{ABIL}_i \mid \text{EXP}]$ does not vary across years for a worker) and we have multiple observations for each worker, we can use "worker fixed effects" ($d1 - d4$) to represent these omitted variables. One can think of $\beta_0 + \beta_2 E[\text{ABIL}_i \mid \text{EXP}]$ as implying a different "intercept" for each worker (for the fitted line through the scatterplot of wage and experience). $d1 \sim d4$ allow the intercept to differ for each worker.