Problem Set 2
Economics 60303
(Due: Friday, February 14, 2014)

Bill Evans
Spring 2014

1. Consider the two-way fixed-effects model $Y_{it} = \alpha + X_{it}\beta + u_i + v_t + \varepsilon_{it}$ where X is a scaler. In a data set with balanced panels and T=2 observations per group, the author estimates the model as a first difference $\Delta Y_i = \theta + \Delta X_i \beta + \Delta\varepsilon_i$. What does the coefficient on θ represent in this case? The researcher then estimates a first difference model with an entire vector of individual fixed effects $\Delta Y_i = \theta + \Delta X_i \beta + \lambda_i + \Delta\varepsilon_i$ and they estimate this model by adding a dummy for n-1 panels to the first difference model, so the equation they estimate is of the form

$$\Delta Y_i = \theta + \Delta X_i \beta + \sum_{i=1}^{n-1} D_i \pi_i + \Delta\varepsilon_i$$

where D_i is defined as in class. In this model, what is being estimated by the parameter $\hat{\pi}_i$?

2. Consider a two way fixed-effects model of the form $y_{it} = \alpha + x_{it}\beta + u_i + \lambda_t + \varepsilon_{it}$ but with unbalanced panels. Let i=1,2,…n and T=t_1, t_2, …t_n. Show that when data is unbalanced, you *cannot* estimate the fixed effect model by de-meaning the data, that is, constructing $\tilde{y}_{it} = y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}$ does not eliminate the fixed effects. HINT: Show that when you construct $\tilde{y}_{it} = y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}$ that the fixed effects u_i and λ_t don't cancel.

3. An author has a balanced panel data set (NT observations) and considers estimating an equation of the form

$$y_{it} = x_{it}\beta + \upsilon_t + \varepsilon_{it}$$

where υ_t is a random year effects where $E[\upsilon_t] = 0$ and $Var(\upsilon_t) = \sigma_\upsilon^2$.

a) Suppose the data is sorted **by year then group** and we can write the equation in matric notation as

$$Y = X\beta + V$$

Where *V* is the composite error and $V_{it} = \upsilon_t + \varepsilon_{it}$. Using Kroeneker products, what is $E[VV']$?

b) How does your answer change if the data is sorted **by group then year**?

5. An author has a balanced panel data set and considers estimating a model of the form

$$y_{it} = x_{it}\beta + v_{it}$$

Where x_{it} is a (kx1) row vector. In this case v_{it} has a three-part error structure

$v_{it} = u_i + \lambda_t + \varepsilon_{it}$ where u_i is a random effect that varies across individuals but is common over time for person i, and λ_t is a random year effect that is common to all groups but is varying randomly across years. Assume $E[u_i] = E[\lambda_t] = E[\varepsilon_{it}] = 0$, $Var(u_i) = \sigma_u^2$, $Var(\lambda_t) = \sigma_\lambda^2, Var(\varepsilon_{it}) = \sigma_\varepsilon^2$ and the errors are random effects such that $\text{cov}(u_i, x_{jit}) = \text{cov}(\lambda_t, x_{jit}) = 0.$

Write the model in matrix form $Y = X\beta + V$ where Y is an NTx1 vector.

a) What is $E[VV']$?

b) What would be the GLS estimate of $\hat{\beta}$? (Just write the equation).

6. Consider a simple difference in difference model where there are two groups (treatment and control, 1 and 0) and two periods (before and after, 1 and 0). Let the equation that measures outcomes be defined by $y_i = \alpha + G_i\gamma + A_i\theta + D_i\beta + \varepsilon_i$ (we have dropped the t subscripts to keep notation easy) where G is the group dummy variable (treatment or not), A is a dummy for after treatment or not, and D is the treatment effect where D=AG. Define $\bar{y}^{00}$ as the mean outcome for the control group in period 0 (prior to treatment) and let $\bar{y}^{01}$ be the mean outcome for the control group after treatment. Let y_i^1 be the outcomes ONLY for the treatment group. Show that a regression of $y_i^1 - [(1-A_i)\bar{y}^{00} + A_i\bar{y}^{01}]$ on a dummy variable for A_i for only the treatment group will produce the difference-in-difference estimate. HINT: write down the definitions of $\bar{y}^{01}$ and $\bar{y}^{00}$.

7. A fellow classmate has written their own code to estimate within, between, and random effects models with balanced panels. As a check on their work, they estimate between, within and random effects models for the one-way effects models with one covariate. The equation that describes the fixed and random effects models is $Y_{it}=X_{it}\beta + u_i + \varepsilon_{it}$. The results are summarized below. You look at the results and tell the student they have a programming error. You are so smart. What tipped you off?

Variable	Between	Fixed	Random
X_{it}	0.01092 (0.00214)	0.03285 (0.00251)	0.03312 (0.00267)

Your classmate fixes their coding error and generates the following results for a different problem. These are correct. What is the Hausman test statistic for the null hypothesis that u_i and X_{it} are uncorrelated? Can you reject or not reject the null? You can calculate this by hand.

Parameter Estimates and Standard Errors

	Between	Fixed	Random
	0.11012	0.04813	0.0635
	(0.04412)	(0.02429)	(0.02137)

Computer Problem

1. In this problem, we are going to examine the roll of measurement error in OLS and fixed-effects models. Load up the data set psid1.dta, construct wages and the natural log of wages (call it wagel) and keep the 1st two years of data (`keep if year<=1984`). The only covariate in these regressions will be the variable "tenure".

   a) Generate the variance of tenure. Call this $\hat{\sigma}_t^2$.

   b) Run an OLS model of wagel on tenure (ignore the panel nature of the data). Then with areg, run the same regression absorbing the individual "id". What are the coefficients and standard errors on tenure in these cases? Call these least-square estimates $\hat{\beta}_{w/out}^{ols}$ and $\hat{\beta}_{w/out}^{fe}$.

   c) Next, you are going to draw random errors, add them to tenure to generate measurement error, estimate the OLS and fixed-effect models, and do this 1000 times, keeping the estimates from every iteration. A sample program that generates random errors for Y and runs the regression of mismeasured Y on a bunch of covariates is called sample_do_loop.do and the data set that accompanies this example is called data_yx. Use this as a template for what you need to do.

   I want the measurement error added to tenure to be from a normal distribution with a mean of 0 and a variance of 4. Given a draw to a normal distribution, you can easily construct a draw from a standard normal distribution with a mean of zero and a variance of 1 by using a draw to a uniform distribution, take the incerse CDF, and multiplying this by 2. Call the measurement error v1 and use the following command to construct the variable:

```
gen v=2*invnorm(uniform())
```

   Add this to tenure 2 and call the new tenure variable with measurement error tenure2

   gen tenure2=tenure+v;

   For each bootstrap iteration, keep four results a) the mean of v, b) the variance of v (call this $\hat{\sigma}_v^2(i)$ c) the OLS estimates with measurement error (call this $\hat{\beta}_{with}^{ols}(i)$), and d) the fixed-effect estimates with measurement error (call this $\hat{\beta}_{with}^{fe}(i)$).

   From the 1000 replications, generate the following values

$$\bar{\hat{\sigma}}_v^2 = \frac{1}{1000}\sum_{i=1}^{1000}\hat{\sigma}_v^2(i)$$
$$\bar{\hat{\beta}}_{with}^{OLS} = \frac{1}{1000}\sum_{i=1}^{1000}\hat{\beta}_{with}^{OLS}(i)$$
$$\bar{\hat{\beta}}_{with}^{fe} = \frac{1}{1000}\sum_{i=1}^{1000}\hat{\beta}_{with}^{fe}(i)$$

d) What are the estimates for $\bar{\hat{\sigma}}_v^2$ $\bar{\hat{\beta}}_{with}^{OLS}$ and $\bar{\hat{\beta}}_{with}^{OLS}$

e) Construct the following measure of the reliability ratio across all 1000 trials:

$$\hat{\theta} = \frac{\hat{\sigma}_t^2}{\bar{\hat{\sigma}}_v^2 + \hat{\sigma}_t^2}\ ?$$

f) Compare the ratio in e) to the ratio $\frac{\bar{\hat{\beta}}_{with}^{OLS}}{\hat{\beta}_{w/out}^{ols}}$. Do these results make sense?

g) Construct the ratio of estimates in the fixed-effects model with the results from part b) $\frac{\bar{\hat{\beta}}_{with}^{fe}}{\hat{\beta}_{w/out}^{fe}}$

h) Devise a method that measures the within-panel correlation in tenure. Using this number and the reliability ratio in e), what should be the attenuation bias associated with measurement error in this case? Compare this to the ratio of estimates from part g).