Economics 30331 Prof. Kasey S. Buckles

Problem Set 4
(due at the beginning of class April 30, 2007)

I. Problems

For each of the following questions, choose the one best answer. Briefly explain your reasoning.

1. Suppose that Vermont has passed a law requiring employers to provide 6 months of paid maternity leave. You are concerned that women's wages will drop in order to pay for this new benefit. You find a data set that samples men and women in Vermont and has information on wages. You pool 2 cross-sections, one from the year before the law took effect and one from the year after and find that the mean wage for various groups is as follows:

	Before	After
Women	$8	$10
Men	$10	$12

Suppose you estimate the following model:

$$wage = \beta_0 + \beta_1 after + \beta_2 women + \beta_3 after{*}women + u,$$

where *after* and *women* are dummy variables for the second period and being a woman respectively. Your estimate of β_3 will be:
a) 2
b) 0
c) -1
d) impossible to say

2. and 3. Suppose a researcher is interested in whether having a lot of college students in a city affects the price of rental housing. Suppose that the true population model is $lrent_{it} = \beta_0 + \beta_1 lpop_{it} + \beta_2 lavginc_{it} + \beta_3 pctstu_{it} + \beta_4 y90_t + a_i + u_{it}$, where *lrent* is the log of the rental price, *lpop* is the log of the city's population, *lavginc* is the log of per capita income, *pctstu* is the student population as a percent of the city population (during the school year) and *y90*=1 if the year is 1990. The researcher uses the fixed effect estimator to obtain the following Stata output:

```
Regression with robust standard errors                Number of obs =     128
                                                      F(  4,    60) =  691.38
                                                      Prob > F      =  0.0000
                                                      R-squared     =  0.9827
                                                      Adj R-squared =  0.9633
                                                      Root MSE      =  .06373
------------------------------------------------------------------------------
             |               Robust
       lrent |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lpop |   .0722458   .0696803     1.04   0.304    -.0671357    .2116272
     lavginc |   .3099605   .0893101     3.47   0.001     .1313138    .4886072
      pctstu |   .0112033    .002936     3.82   0.000     .0053305    .0170761
         y90 |   .3855214   .0487188     7.91   0.000     .2880693    .4829735
       _cons |   1.409384   1.162338     1.21   0.230    -.9156381    3.734405
-------------+----------------------------------------------------------------
        city |   absorbed                                     (64 categories)
------------------------------------------------------------------------------
```

2. Based on this, we can conclude that
a) the researcher thinks the unobserved fixed effect is likely correlated with the observed x variables
b) the researcher thinks these unobserved fixed effects cause serial correlation in the errors
c) the researcher thinks that the error term is heteroskedastic
d) all of the above

3. Suppose another researcher had the same data and regressed $\Delta lrent_i$ on $\Delta lpop_i$, $\Delta lavginc_i$ and $\Delta pctstu_{it}$. We can say for sure that
a) these estimates would be different, but the implications would be very similar to the previous model
b) the estimated constant term would be .3855214
c) the estimated constant term would be 1.409384
d) none of the above

4. Which of the following are true about time-series estimation?
 a) As with cross-section estimation, we can assume the observations are independent
 b) A trending variable cannot be used as the dependent variable in multiple regression
 c) Serial correlation in the error terms causes biased estimates.
 d) None of the above are true.

Part II. Stata Problems. See "Important Things to Know about Stata" on our website for more about installing and using Stata.

5. For this problem, you will use the Stata data set **wage2.dta** that is on our web page.

a) Regress lwage on educ, exper, tenure, IQ, married, black, south and urban. Based on this regression, how large are the returns to education? Why might this be an overestimate? Why might this be an underestimate?

b) Potential instruments for for *educ* are *sibs*, *meduc* and *feduc*. Run the first-stage regression of 2SLS to check on whether these variables are indeed related. What attributes are needed in an instrument?

c) Estimate the model from a) using IV estimation, where the instruments for for *educ* are *sibs*, *meduc* and *feduc*, as discussed above. Now how large are the returns to education? What might this indicate about the most important problem with the estimate from part a?

> *Stata Notes*: While 2SLS estimates can be calculated by hand, the standard errors will not be correct. Thus, it is preferable to let Stata do the work. We just need to know which variables are exogeneous and which are endogenous. The exogenous variables and the instruments for the endogenous variables then appear in parentheses after the reg command. To regress y on x1, x2 and x3, where z1 is an instrument for x1, and z2a and z2b are instruments for x2, and thus x3 is assumed exogenous, just type **reg y x1 x2 x3 (z1 z2a z2b x3)**

6. This problem uses **murder.dta** which contains state-level data on murder rates in 1987, 1990 and 1993. It assumes a researcher is interested in whether the death penalty serves as a deterrent. Take a close look at what variables are in the data set and what their means are before getting started.

a) Suppose that a researcher is modeling the murder rate simply as a function of the unemployment rate and the number of executions in the past 3 years. Based on this simple model, carry out a Hausman test to determine whether a fixed effect model is preferred. What is the null hypothesis being tested?

b) Estimate this model using a fixed-effects approach, adjusting the standard errors for possible heteroskedasticity.

c) Based on the above, what would you conclude about the possible deterrent effects of the death penalty?

7. This problem uses the Stata data set **intdef.dta** that includes data on inflation, interest rates and the deficit. It assumes a researcher is interested in the effect of interest rates and deficits on inflation. Be sure to look carefully at the variable labels and means before starting.

a) Estimate a finite distributed lag model (of order 1) of the effect of inflation **and** the deficit on the interest rate. What is the impact propensity for inflation? What about for the deficit? Explain your answer.

b) What is the long-run propensity for inflation? What about for the deficit? In each case, test the null hypothesis for whether the long-run propensity is significantly different from zero. Discuss your results.

c) Do a test for AR(1) serial correlation in this model. Do you reject or fail to reject the null hypothesis of no serial correlation?

d) Re-estimate the equation from part 1 but with standard errors robust to AR(1) serial correlation. Compare these results to those obtained in part a).