## Part II: 75 points

This part of the test asks a series of questions about the relationship between *dthrte*, traffic death rates (number of traffic deaths per million miles driven), *unem*, unemployment rates, and *admn*, a dummy variable that equals 1 if a State in a given year has administrative laws such as driver licenses can be suspended for drunk driving. Data was gathered for these variables for the 48 contiguous States of the United States (Hawaii, Alaska and Washington D.C. are not included) for the years 1980 and 2004. The average death rate across all states was 3.55 in 1980 and was 1.51 in 2004.

Suppose we are interested in the following regression relationship

$$dthrte_{it} = \beta_0 + \delta_0 d2_t + \beta_1 admn_{it} + \beta_2 unem_{it} + a_i + u_{it},$$

where $t = 1$ is 1980 and $t = 2$ is 2004. The variable $d2_t$ is a dummy variable that equals 1 if $t = 2$ (2004). The total error depends on the State specific heterogeneity, $a_i$, and an idiosyncratic error, $u_{it}$. We would expect $a_i$ to be correlated with $admn_{it}$ because states with higher than normal traffic death rates may be under pressure to pass drunk driving laws. You can assume that $u_{it}$ is not correlated with $admn_{it}$ and $unem_{it}$.

Consider four options for estimating the parameters $\beta_1$ and $\beta_2$. For the first two options we estimate separate regressions for 1980 ($t = 1$) and 2004 ($t = 2$) using OLS:

$$dthrte_{i1} = \beta_0 + \beta_1 admn_{i1} + \beta_2 unem_{i1} + a_i + u_{i1},$$

$$dthrte_{i2} = \alpha_0 + \beta_1 admn_{i2} + \beta_2 unem_{i2} + a_i + u_{i2}.$$

For the third option we can pool the data from both years and estimate the panel model

$$dthrte_{it} = \beta_0 + \delta_0 d2_t + \beta_1 admn_{it} + \beta_2 unem_{it} + a_i + u_{it},$$

using OLS (i.e. the Pooled-OLS estimators). Finally, we can apply the first difference transformation to the panel model and estimate the transformed model

$$\Delta dthrte_i = \delta_0 + \beta_1 \Delta admn_i + \beta_2 \Delta unem_i + \Delta u_i.$$

by OLS (i.e. the First-Differences estimators).

a) What are the interpretations of the $\delta_0$, $\beta_1$ and $\beta_2$ parameters? **(6 points)** If you used First-Differences to estimate these parameters, would this change the interpretation? Why or why not? **(2 points)**

The four estimation methods yielded the following results given in Table 1. Heteroskedasticity robust standard errors are given in parentheses.

**Table 1: Estimates for Death Rate Model Using the Years 1980 and 2004**

| | 1980 only | 2004 only | Pooled OLS | First Differences |
|---|---|---|---|---|
| $\widehat{\beta}_1$ (*admn*) | 1.56 | -.0065 | .159 | -.308 |
| | (.833) | (.154) | (.224) | (.238) |
| $\widehat{\beta}_2$ (*unem*) | -.070 | .028 | -.019 | .021 |
| | (.077) | (.062) | (.057) | (.071) |
| $\widehat{\delta}_0$ (*d*2) | NA | NA | -2.19 | -1.75 |
| | | | (.222) | (.265) |
| *obs* | 48 | 48 | 96 | 48 |

b) Provide an explanation for why the estimates of $\beta_1$ are so different across the four estimation methods. **(6 points)**

c) Which estimator of $\beta_1$ do you prefer and why? **(4 points)** Is your preferred estimator of $\beta_1$ statistically significant at the 10% level? Why or why not? **(2 points)** Is it practically significant? Why or why not? **(2 points)**

d) Using the First Differences estimator, test the null hypothesis that mean traffic death rates did NOT fall from 1980 to 2004 (holding *admn* and *unemp* constant) against the hypothesis that mean traffic death rates did fall. Write down the null and alternative hypotheses in terms of the relevant parameter and carry out your test at the 5% level. **(6 points)** What do you conclude? **(1 point)**

e) From the Table we see that First Differences is based on only one year's worth of data ($obs$ = 48). Someone suggests that Fixed-Effects is a better estimation option than First Differences because there is unlikely to be much, if any, correlation between $u_{it}$ at years 1980 and 2004 because shocks to traffic death do not last for 25 years. Suppose it is true that $u_{i1}$ (1980) is uncorrelated with $u_{i2}$ (2004). Can Fixed-Effects be better than First Differences? Why or why not? **(6 points)**

An intern was hired to collect data for all the years between 1980 and 2004 increasing the sample size of the panel from 96 to 1,200 observations. We now have $T = 25$ rather than $T = 2$. The model of interest is the same except that dummy variables are included for the additional time periods:

$$dthrte_{it} = \beta_0 + \delta_2 d2_t + \delta_3 d3_t + ... + \delta_{25} d25_t + \beta_1 admn_{it} + \beta_2 unem_{it} + a_i + u_{it}, \qquad (1)$$

Model (1) was estimated using Pooled-OLS, First Differences, Fixed-Effects and Random Effects. The results are in given in the following table. Time dummy parameter estimates are only reported for every 5th year. The estimated value of $\theta$ used for the random effects transformation was $\widehat{\theta} = .749$. All standard errors are robust to heteroskedasticity and correlation across time (serial correlation).

**Table 2: Estimates for Death Rate Model Using the Years 1980 through 2004**

| | Pooled OLS | First Differences | Fixed Effects | Random Effects |
|---|---|---|---|---|
| $\widehat{\beta}_1$ $(admn)$ | .031 | -.115 | -.140 | -.136 |
| | (.114) | (.039) | (.058) | (.056) |
| $\widehat{\beta}_2$ $(unem)$ | .080 | -.023 | -.043 | -.039 |
| | (.030) | (.010) | (.015) | (.015) |
| $\widehat{\delta}_5$ $(d5, 1984)$ | -1.03 | -.864 | -.848 | -.851 |
| | (.092) | (.068) | (.066) | (.069) |
| $\widehat{\delta}_{10}$ $(d10, 1989)$ | -1.22 | -1.33 | -1.35 | -1.35 |
| | (.110) | (.085) | (.086) | (.086) |
| $\widehat{\delta}_{15}$ $(d15, 1994)$ | -1.70 | -1.73 | -1.74 | -1.74 |
| | (.130) | (.089) | (.092) | (.088) |
| $\widehat{\delta}_{20}$ $(d20, 1999)$ | -1.72 | -1.88 | -1.92 | -1.92 |
| | (.149) | (.093) | (.101) | (.100) |
| $\widehat{\delta}_{25}$ $(d25, 2004)$ | -1.91 | -1.98 | -2.00 | -2.00 |
| | (.140) | (.093) | (.102) | (.101) |
| *obs* | 1200 | 1152 | 1200 | 1200 |

f) Assuming that $cov(admn_{it}, a_i) \neq 0$, which estimation methods of model (1) could be LUE? Why? **(3 points)** Which estimation methods are biased? Why? **(3 points)**

g) For each of the potentially LUE estimators, under what assumptions will that estimator be BLUE? **(4 points)**

h) The Pooled-OLS estimator of $\beta_1$ has the opposite sign as the First-Difference and Fixed-Effects estimators. Provide a theoretical explanation for this fact. **(4 points)** The Random Effects estimator of $\beta_1$ is very similar to Fixed-Effects. Are you surprised by this? Why or why not? **(2 points)**

i) Calculate 95% confidence intervals for $\beta_1$ using the First Difference and Fixed-Effects estimators. **(4 points)** Can you reject the null hypothesis that $\beta_1 = 0$ with either confidence interval? **(2 points)** Which estimator would you use for measuring the impact of administrative laws on traffic death rates? Why? **(2 points)**

j) Go back to Table 1 that only uses the 1980 and 2004 data and compute a 95% confidence interval for $\beta_1$ using the First Difference estimator. **(2 points)** You should find that this confidence interval is substantially wider than the confidence intervals using all the data from 1980 to 2004. What accounts for this difference in confidence interval widths? **(2 points)** What do you conclude about using just the 1980 and 2004 data compared to using all the years from 1980 to 2004? **(2 points)**

k) Using the First Differences estimator from Table 2 (all years from 1980 to 2004), test the null hypothesis that mean traffic death rates did NOT fall from 1980 to 2004 (holding *admn* and *unemp* constant) against the hypothesis that mean traffic death rates did fall. Write down the null and alternative hypotheses in terms of the relevant parameter and carry out your test at the 5% level. **(2 points)** How do your results compare to what you found in part (d)? **(2 points)** Provide an explanation for any differences you see. **(2 points)**

l) Using the Fixed-Effects estimated parameters, holding *admn* and *unem* constant, over what 5 year period did traffic death rates fall more: 1984 to 1989, 1989 to 1994, 1994 to 1999, or 1999 to 2004? Provide details on your calculations. **(4 points)**