**Econ 513: Practice of Econometrics**

**Final exam, December 8, 2016, Answers**

*Problem 1.*

(i) For all $i$ the dummies $d_{ij}$ have a sum equal to 1. Therefore we have a multicollinearity problem, because the regressor for the intercept that is 1 is equal to the sum of the dummies $d_{ij}$.

(ii) If we consider $\alpha_i$ as a random effect it is uncorrelated with $x_{i1}, x_{i2}, x_{i3}$. If it is a fixed effect it can be correlated with $x_{i1}, x_{i2}, x_{i3}$.

(iii) The decision on the number of employees is taken by management. This implies that the decision will depend among other things on the quality of the management $\alpha_i$, so that the number of employees is correlated with $\alpha_i$. The sign of the correlation could be negative if good management means higher sales and fewer employees are needed for a level of sales. We need not know the sign of the correlation to make the fixed effect assumption more credible in this case.

(iv) FD model

$$\Delta y_{it} = \beta \Delta x_{it} + \Delta \varepsilon_{it},\ t = 2, 3$$

The assumption is (I express the lack of relation by a 0 correlation assumption) $E(\Delta x_{it} \Delta \varepsilon_{it}) = 0,\ t = 2, 3$ and a sufficient condition for that is

$$E(x_{it}\varepsilon_{it}) = 0,\ t = 1, 2, 3 \quad E(x_{it}\varepsilon_{i,t-1}) = 0,\ t = 2, 3 \quad E(x_{i,t-1}\varepsilon_{it}) = 0,\ t = 2, 3$$

or

$$E(x_{it}\varepsilon_{it}) = 0,\ t = 1, 2, 3 \quad E(x_{it}\varepsilon_{i,t-1}) = 0,\ t = 2, 3 \quad E(x_{it}\varepsilon_{i,t+1}) = 0,\ t = 1, 2$$

(v) This implies that the regression (1) has a linear time trend $\gamma \cdot t$.

(vi)

$$y_{i3} = \gamma y_{i2} + \beta x_{i3} + \alpha_i + \varepsilon_{i3}$$

$$y_{i2} = \gamma y_{i1} + \beta x_{i2} + \alpha_i + \varepsilon_{i2}$$

For period 1 the lagged dependent variable is for period o for which we do not have observations.

(vii)

$$\Delta y_{i3} = \gamma \Delta y_{i2} + \beta \Delta x_{i3} + \Delta \varepsilon_{i3}$$

We require that the error $\varepsilon_{i3} - \varepsilon_{i2}$ is uncorrelated with $y_{i2} - y_{i1}$ and $x_{i3} - x_{i2}$ and a sufficient condition is

$$E(x_{i3}\varepsilon_{i3}) = 0 \quad E(x_{i3}\varepsilon_{i2}) = 0 \quad E(x_{i2}\varepsilon_{i3}) = 0 \quad E(x_{i2}\varepsilon_{i2}) = 0$$

and

$$E(y_{i2}\varepsilon_{i3}) = 0 \quad E(y_{i2}\varepsilon_{i2}) = 0 \quad E(y_{i1}\varepsilon_{i3}) = 0 \quad E(y_{i1}\varepsilon_{i2}) = 0$$

(viii)(20) Because

$$y_{i2} = \gamma y_{i1} + \beta x_{i2} + \alpha_i + \varepsilon_{i2}$$

we have that $y_{i2}$ and $\varepsilon_{i2}$ are correlated so that

$$E(y_{i2}\varepsilon_{i2}) \neq 0$$

To deal with this correlation we can use IV with $y_{i1}$ and $x_{i1}$ potential instruments for $\Delta y_{i2}$.

*Problem 2.* In a study of the effect of financial aid on graduation from a college let the outcome variable be $Y$ that is 1 if the student graduates and 0 if s/he drops out. The dummy $D$ is 1 if the student receives financial aid and 0 if not. The variable $Z$ is a score (for instance high school GPA) that measures the performance of the student in high school. The data are a sample $Y_i, D_i, Z_i, i = 1, \ldots, N$ of students who were admitted to the same college.

(i) In the regression

$$Y = \alpha + \beta D + \varepsilon \tag{1}$$

$D$ and $\varepsilon$ are likely to be positively correlated and this will bias the OLS estimate upward.

(ii) Although $Z$ and $D$ are possibly strongly correlated the problem is that $Z$ and $\varepsilon$ are also likely to be correlated, because $Z$ has a direct effect on $Y$ even after accounting for $D$.

(iii) No. Even if $D$ is a function of $Z$ only then $Z$ and $\varepsilon$ are still correlated so that $D$ is still correlated with the error term.

(iv)The Regression Discontinuity (RD) estimator would apply. For this estimator we would estimate $E(Y|Z)$ by a local linear regression using data just above the cut-off and estimate

$$\lim_{z \downarrow c} E(Y|Z = z)$$

and we would estimate $E(Y|Z)$ by a local linear regression using data just below the cut-off and estimate

$$\lim_{z \uparrow c} E(Y|Z = z)$$

The RD estimator is

$$\lim_{z \downarrow c} E(Y|Z = z) - \lim_{z \uparrow c} E(Y|Z = z)$$

(v) The assumption is that $E(Y_0|Z = z)$ is continuous in $z$ and $E(Y_1|Z = z)$ is continuous in $z$. Alternatively we can require that for $\delta$ small $E(\varepsilon|Z = z - \delta) = E(\varepsilon|Z = z + \delta)$ so that

$$E(Y|Z = z+\delta) = \alpha+\beta+E(\varepsilon|Z = z+\delta) \quad E(Y|Z = z-\delta) = \alpha+E(\varepsilon|Z = z+\delta)$$

so that under the assumption above

$$E(Y|Z = z + \delta) - E(Y|Z = z - \delta) = \beta.$$

(vi) In that case there would be more students with scores above that just below the cut-off and this would show up in the density of the distribution of the scores as a dis

(vii) (20)

$$E(Y|Z = z+\delta)-E(Y|Z = z-\delta) = \beta(E(D|Z = z+\delta)-E(D|Z = z-\delta))+E(\varepsilon|Z = z+\delta)-E(\varepsilon|Z = z-\delta)$$

Now

$$E(\varepsilon|Z = z + \delta) - E(\varepsilon|Z = z - \delta) = 0$$

because the students just below and above the cut-off are similar (except for the receipt of financial aid). Therefore

$$\beta = \frac{E(Y|Z = z + \delta) - E(Y|Z = z - \delta)}{E(D|Z = z + \delta) - E(D|Z = z - \delta)}$$

and we use the procedure in (iv) to estimate numerator and denominator.