Homework #4

Due in class 12/2/16

1) (**CRE vs Fixed Effects)**

Consider the model

$$Y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it}$$

where α_i is a fixed person effect, X_{it} is a scalar regressor, and ε_{it} is a strictly exogenous time varying error. Assume the data are balanced, with T observations on each individual i in the sample.

a) Prove that the least squares dummy variable (LSDV) estimator of this model is consistent for β.

b) Suppose we invoke the following correlated random effects assumption: $\alpha_i = a_0 + a_1 \overline{X}_i + v_i$, where $\overline{X}_i = \frac{1}{T} \sum_{t=1}^{T} X_{it}$ and $E\left[v_i | X_{i1}, ..., X_{iT}\right] = 0$. Then we may write

$$Y_{it} = a_0 + a_1 \overline{X}_i + \beta X_{it} + u_{it}, \tag{1}$$

where $u_{it} = v_i + \varepsilon_{it}$. Show that least squares applied to the above equation yields an estimate of β numerically equivalent to the LSDV estimator.

c) Describe a procedure for testing for zero correlation between the person effects and X_{it}.

d) Propose estimators for $Var\left(v_i\right) = E\left[\left(\alpha_i - a_0 - a_1 \bar{X}_i\right)^2\right]$ and for $Var\left(\alpha_i\right)$.

2) (**Event Study**) Download the Stata file curfews_class.dta from Bcourses. This file contains data on the number of arrests of youth in a panel of cities enacting juvenile curfew laws in different years.

a) Construct a variable E_{it} that equals one in the year that a city enacts a curfew law (e.g. when year==enacted).

b) Run an event study of log juvenile arrests on curfew enactment with city and year effects (hint: using the "tsset" command will make this easier since you can then use the "L." operator to construct leads and lags). Be sure to "bin up" the endpoints and to cluster the standard errors by city. If you bin up, you will need to leave out one event dummy since they sum to a constant (everyone is eventually treated). As a normalization, leave out E_{it-1} so that all effects can be measured relative to the period before enactment.

c) Make a plot of the event study coefficients with confidence intervals. What do you conclude about the impact of the program on arrests?

d) Try changing the number of leads and lags in your event study. How do the coefficients change?

e) Try adding city specific trends to the event study as controls. How do the coefficients change?

3) (**Dynamic Panel**) Having looked at the event study you decide to try parameterizing the treatment effect to obey a simple partial adjustment model of the form:

$$\begin{aligned} Y_{it} &= \alpha_i + \gamma_t + \delta Y_{it-1} + \beta_0 E_{it} + \beta_1 E_{it-1} + \varepsilon_{it} \qquad (2) \\ & \varepsilon_{it} \overset{iid}{\sim} N\left(0, \sigma^2\right) \end{aligned}$$

a) Take first differences of the above model in order to eliminate the fixed effect and run ΔY_{it} on $\Delta Y_{it-1}, \Delta E_{it}, \Delta E_{it-1}$, and year dummies. Interpret your results.

b) Arellano and Bond (1991) suggest instrumenting for ΔY_{it-1}. Try using Y_{it-2} as an instrument. Now try using ΔY_{it-2} as an instrument. Now use them both. How do the results change?

c) Compute the overidentification statistic arising from the use of both Y_{it-2} and ΔY_{it-2} as instruments using the ivreg2 plugin (Stata's ivregress command has a bug in its calculation of cluster robust J tests). Interpret your results.

d) Compare the causal responses to treatment implied by the dynamic panel model to those from the event study.

e) Which approach to analyzing the data do you find most convincing? Why?

4) (**CRE Event Study**) The event study design relies on potentially strong parametric assumptions. One approach to testing those assumptions is to use a Chamberlain (1982) style CRE approach. Consider the projection of the time invariant city effects α_i on the city curfew enactment dates e_i:

$$\alpha_i = \eta_0 + \sum_{t=t_0}^{T} \eta_t 1\left[e_i = t\right] + v_i$$

where v_i is an uncorrelated random effect. We therefore have the following restricted reduced form for log arrests:

$$Y_{it} = \eta_0 + \gamma_t + \sum_{k=0}^{T-t_0} \delta_k 1\left[e_i = t-k\right] + \sum_{l=t_0}^{T} \eta_l 1\left[e_i = l\right] + v_i + \varepsilon_{it} \tag{3}$$

where the $\{\delta_k\}$ give the dynamic causal effects of curfew laws and the $\{\eta_l\}$ represent endogeneity.

a) Reshape the data into wide format so that there is a separate lnarrests variable for each year. To keep the exercise manageable restrict the sample to years 1984-1990.

b) Over the sample period in question, there are only 5 event dates: $\{85, 87, 88, 89, 90\}$. Use Stata's "mvreg" command to estimate seven unrestricted cross sectional regressions of the form:

$$Y_{it} = \tilde{\pi}_t + \sum_{k \in \{85,87,88,89,90\}} \pi_{k,t} 1\left[e_i = k\right] + r_{it} \text{ for } t \in \{84, 85, 86, 87, 88, 89, 90\} \tag{4}$$

where the $\{\pi_{k,t}\}$ are the unrestricted reduced form regression coefficients and the $\{\tilde{\pi}_t\}$ are reduced form year specific intercepts.

c) Use Stata's "test, accum" post-estimation command to test the linear restrictions on the $\{\pi_{k,t}\}$ implied by the model in (3). For example, one restriction is that $\pi_{k,t} = \pi_{k,t'}\ \forall (t, t') < k$. Enumerate all of the restrictions and test them jointly. Can you reject the model? What important inferential assumptions are you using in constructing your p-values?

d) Use Stata's "constraint" and "sureg" commands to estimate (4) subject to the restrictions in (3). Be sure to use the small sample degrees of freedom adjustment. Then use Stata's "lincom" command to recover estimates (and standard errors) of $\{\delta_0, \delta_1, \delta_2, \delta_3, \delta_4, \delta_5\}$ from the constrained $\{\pi_{k,t}\}$. How do these estimates compare with those from the event study in question 2)?