New York University
Applied Statistics and Econometrics II
Spring 2016

Problem Set 2

(I) Use the data in *crime.txt* for the years 1972 and 1978 for a two-year panel data analysis. The model is a simple distributed lag model:

$$\log(crimerate_{it}) = \theta_0 + \theta_1 d78_t + \beta_1 clrprc_{i,t-1} + \beta_2 clrprc_{i,t-2} + c_i + u_{it}$$

The variable *clrprc* is the clear-up percentage (the percentage of crimes solved). The data are stored for two years, with the needed lags given as variables for each year.

1. First estimate this equation using a pooled OLS analysis. Comment on the deterrent effect of the clear-up percentage, including interpreting the size of the coefficients. Test for serial correlation in the composite error v_{it} assuming strict exogeneity.
2. Estimate the equation by fixed effects, and compare the estimates with the pooled OLS estimates. Is there any reason to test for serial correlation? Optional: obtain heteroscedasticity - robust standard errors for the FE estimates.
3. Estimate the model using random effects. Indicate how to test for random effects and show the result of your test.
4. Using FE analysis, test the hypothesis $H_0 : \beta_1 = \beta_2$. What do you conclude? If the hypothesis is not rejected, what would be a more parsimonious model? Estimate this model.

(II) Suppose that you observe a sequence of observations $y = (y_1, \dots, y_n)$. The variance of the process σ^2 is known. Your prior beliefs for the mean are:

$$\mathbb{P}(\theta) = \frac{1}{\sqrt{2\pi\tau_0^2}} \exp\left(-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2\right) \tag{1}$$

And the data distribution conditional on the parameters is normal:

$$\mathbb{P}(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y - \theta)^2\right) \tag{2}$$

Using Bayes's law, calculate the posterior kernel.

(III) For $j = 1, 2$ suppose that:

$$(y_{j1}, \dots, y_{jn}|\mu_j, \sigma_j^2) \sim iidN(\mu_j, \sigma_j^2)$$

$$\mathbb{P}(\mu_j, \sigma_j^2) \propto \sigma_j^{-2}$$

and $\perp \{(\mu_1, \sigma_1^2), (\mu_2, \sigma_2^2)\}$ in the priors. Show that the posterior distribution of $(s_1^2/s_2^2)/(\sigma_1^2/\sigma_2^2)$ is $F(n_1 - 1, n_2 - 1)$.

(IV) Let us consider application 1 in chapter 16 of Greene. We investigate the mix of male and female children in families. K_i is the family size (number of children), and F_i the number of female children in the family.

Suppose that:

$$\mathbb{P}(F_i|K_i,\theta) = \begin{pmatrix} K_i \\ F_i \end{pmatrix} \theta^{F_i}(1-\theta)^{K_i-F_i} \tag{3}$$

We look for an estimate of θ. The prior over θ is:

$$\mathbb{P}(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$$

Use the sample of observations in the book to obtain:

1. The maximum likelihood estimate of θ from a frequentist perspective.
2. The posterior density of θ given the data conditional on a and b.
3. Compute the posterior mean in a 3-d plot for different values of a and b.