Problem Set 2

Prof. Conlon

Due: 3/1/19

Packages to Install

The packages used this week are

- plm (Panel Data Package)
- estimatr (Tidyverse version of lm function)

Problem 1 (Analytical Exercise)

Consider the estimation of the individual effects model:

$$y_{it} = x'_{it}\beta + \alpha_i + \epsilon_{it},\, \mathbb{E}\big[\epsilon|x_{it}, \alpha_i\big] = 0$$

where $i = \{1, ..., n\}$ and $t = \{1, ..., T\}$.
This exercises ask you to relate the (random effects) GLS estimator $\hat{\beta}_{GLS} = (X'_* X_*)^{-1} X'_* y_*$ to the "within" (fixed-effects) estimator $\hat{\beta}_{FE} = (\dot{X}'\dot{X})\dot{X}'\dot{y}$ and the "between" estimator $\hat{\beta}_{BW} = (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{y}$ where $w = \{x, y\}$:

$$\begin{aligned}
\bar{w}_i &:= \frac{1}{T}\sum_{i=1}^{T} w_i \\
\dot{w}_i &:= w_{it} - \bar{w}_i \\
w_{it,*} &:= w_{it} - (1-\lambda)\bar{w}_i \\
\lambda^2 &= \frac{Var(\epsilon)}{T\,Var(\alpha_i) + Var(\epsilon_{it})}
\end{aligned}$$

1. Express the GLS estimator in terms of $\bar{X}$, $\dot{X}$, $\bar{y}$, $\dot{y}$, λ, and T.
2. Show that there is a matrix R depending on $\bar{X}$, $\dot{X}$, λ and T such that the GLS estimator is a weighted average of the "within" and "between" estimators:
$$\hat{\beta}_{GLS} = R\hat{\beta}_{FE} + (I - R)\hat{B}_W$$
3. What happens to the relative weights on the "within" and "between" estimators as we increase the sample size, i.e. $T \to \infty$?
4. Suppose that the random effects assumption $\mathbb{E}\big[\alpha_i|x_{i1,...,x_{iT}}\big] = 0$ does not hold. Characterize the bias of the estimators $\hat{\beta}_{FE}$, $\hat{\beta}_W$. (Note: An estimator $\hat{\beta}$ is unbiased if $\mathbb{E}\big[\hat{\beta}\big] = \beta$)
5. Use your result from (d) to give a formula for the bias of our random effects estimator $\hat{\beta}_{GLS}$. What happens to the bias as $T \to \infty$.

Problem 2 (Analytical Exercise)

Assume we have the following dynamic panel model (with the same indices from Question 1):

$$y_{i,t} = \gamma y_{i,t-1} + \alpha_i + \epsilon_{i,t},$$

We have the following assumptions:

$$
\begin{aligned}
|\gamma| &< 1 \\
y_{i,0} &\text{ is known (i.e. non-random)} \\
\mathbb{E}_T\left[\epsilon_{i,t}|y_{i,t-1},...,y_{i,0},\alpha_i\right] &= 0 \text{ (Sequential Exogeneity)}
\end{aligned}
$$

Our goal is to consistently estimate γ.

1. Show that the first-difference estimator of γ is unbiased if $\alpha_i = 0$.
2. Is the first difference estimator of γ still unbiased if $\alpha_i \neq 0$?
   (Hint: Try showing that $cov(\epsilon_{i,t}, y_{i,t+1}) \neq 0$. Why is this condition useful?)
3. Let $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$, what is $cov(\Delta\epsilon_{i,t}, y_{i,t-2})$? What is $cov(\Delta y_{i,t}, y_{i,t-2})$?
4. Using your answer from part (4), propose a strategy to consistently estimate γ?

Problem 3 (Analytical Exercise)

You will still be working with the same model from the previous question. We now want to investigate the properties of the fixed effects (“within”) estimator.

1. Derive the following formula for the (population) fixed effects ("within") estimator β_{FE} in terms of $\bar{y}_i$, $\bar{y}_{i,-1}$, $y_{i,t-1}$, $y_{i,t}$ where:

$$
\begin{aligned}
\beta_{FE} &= \mathbb{E}_T\left[\left(y_{i,t-1}-\bar{y}_{i,-1}\right)^2\right]^{-1}\mathbb{E}_T\left[\left(y_{i,t-1}-\hat{y}_{i,-1}\right)\left(y_{i,t}-\bar{y}_i\right)\right] \\
\bar{y}_i &= \frac{1}{T}\sum_{t=1}^{T} y_{i,t} \\
\bar{y}_{i,-1} &= \frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}
\end{aligned}
$$

   (Note: $\mathbb{E}_T\left[\,\cdot\,\right]$ is an expectation with respect to some probability measure (think distribution) fixing our sample size T)

2. Show that the potential bias of the FE estimator is:

$$
\hat{\beta}_{FE} - \beta = \mathbb{E}_T\left[\left(y_{i,t-1}-\bar{y}_{i,-1}\right)^2\right]^{-1}\mathbb{E}_T\left[\left(y_{i,t-1}-\bar{y}_{i,-1}\right)\left(\epsilon_{i,t}-\bar{\epsilon}_i\right)\right]
$$

3. Show that the numerator has the following form:

$$
\mathbb{E}_T\left[\left(y_{i,t-1}-\bar{y}_{i,-1}\right)\left(\epsilon_{i,t}-\bar{\epsilon}_i\right)\right] = -\mathbb{E}_T\left[\bar{y}_{i,-1}\bar{\epsilon}_i\right]
$$

   (Hint: Factor out the expression into four terms, work with each term individually)

4. Why do we expect the numerator not to be equal to 0?
   (No need for rigourous calculations)
5. Using backward substitution:

$$
\begin{aligned}
y_{i,t} &= \gamma y_{i,t-1} + \alpha_i + \epsilon_{i,t} \\
y_{i,t} &= \gamma\left(\gamma y_{i,t-2} + \alpha_i + \epsilon_{i,t-1}\right) + \epsilon_{i,t} \\
&\vdots
\end{aligned}
$$

Derive the following form for the numerator:

$$-\mathbb{E}_T\left[\bar{y}_{i,-1}\bar{\epsilon}_i\right] = -\frac{\mathbb{E}_T\left[\epsilon_{i,t}^2\right]}{T^2}\frac{\left(T - T\gamma - 1 + \gamma^T\right)}{\left(1-\gamma^2\right)}$$

(Hint: Remember the following formula: if $|\gamma| < 1$ then $\sum_{t=0}^{T}\gamma^t b = b\left(\frac{1-\gamma^{T+1}}{1-\gamma}\right))$

6. Once again, using backward substitution derive the following form for the denominator:

$$\mathbb{E}_T\left[\left(y_{i,t-1} - \bar{y}_{i,-1}\right)^2\right] = \frac{\mathbb{E}_T\left[\epsilon_{i,t}^2\right]}{1-\gamma^2}\left(1 - \frac{1}{T} - \frac{2\gamma}{(1-\gamma)^2}\frac{(T - T\gamma - 1 + \gamma T)}{T^2}\right)$$

7. What happens to the bias for small T? For $T \to \infty$?
(Note: You need to only reduce the algebra to something you can make a statement about the bias.)

Problem 4 (Coding Exercise)

We observe N observations of the random variable X_i where each X_i is drawn from the Weibull distribution:

$$X_i \sim W(\gamma)$$

The probability density function for the Weibull is the following:

$$f(x;\gamma) = \gamma x^{\gamma-1}\exp(-(x^{\gamma})) \;\; ; x \geq 0, \gamma > 0$$

1. Assume our N observations are independent and identically distributed, what is the log-likelihood function?
2. Calculate the gradient (or first derivative) of your log-likelihood function.
3. In R, I want you to write a function called mle_weibull that takes two arguments (X, γ), where X is a vector of data and γ is a scalar. The function returns the value of the log-likelihood function you derived in the last part.
4. Optimiziation routines can either be given a first derivative (or gradient) or the optimization routines calculate numerical derivatives. We will be using the R function *optim*, which accepts the first derivative as an argument *gr*.
   a. We first want you to run *optim* without supplying a first derivative (leaving gr out of the function). Note, to run optim you will need to supply your data X as an additional parameter at the end of the function. We have provided you with simulated data in the file 'prob_4_simulation.rda' located in the data folder.
   b. We now want you to create a new function called gradient, which takes the same two arguments as your likelihood function. Now calculate the MLE using optim with the gradient.
   c. Compare both the number of iterations until convergence and your estimated γ values from both runs.