Problem Set 2

Prasanna Parasurama

Due: 3/1/19

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.

   Let M be the transformation matrix such that $MX = \dot{X}$
   Let P be the transformation matrix such that $PX = \bar{X}$

$$\hat{\beta}_{GLS} = [X'(M + \lambda P)X]^{-1}[X'(M + \lambda P)y] \tag{1}$$

   Consider $X'(M + \lambda P)X$. Note that M and P are idempotent. That is, $M^2 = M$ and $P^2 = P$.

$$\begin{aligned}
X'(M + \lambda P)X &= X'(MM + \lambda PP)X &(2)\\
&= X'(MMX + \lambda PPX) &(3)\\
&= X'(M\dot{X} + \lambda P\bar{X}) &(4)\\
&= (\dot{X}'\dot{X} + \lambda\bar{X}'\bar{X}) &(5)
\end{aligned}$$

   Likewise:

$$X'(M + \lambda P)y = (\dot{X}'\dot{y} + \lambda\bar{X}'\bar{y}) \tag{6}$$

   Plugging these back into (1):

$$\hat{\beta}_{GLS} = [(\dot{X}'\dot{X} + \lambda\bar{X}'\bar{X})]^{-1}(\dot{X}'\dot{y} + \lambda\bar{X}'\bar{y}) \tag{7}$$

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:

$$\begin{aligned}\hat{\beta}_{GLS} &= R\hat{\beta}_{FE} + (I - R)\hat{\beta}_{BW} \\ &= R\hat{\beta}_{FE} + \hat{\beta}_{BW} - R\hat{\beta}_{W}\end{aligned}$$

where,

$$\begin{aligned}R &= [\dot{X}'\dot{X} + \lambda\bar{X}'\bar{X}]^{-1}\dot{X}'\dot{X} \\ \hat{\beta}_{FE} &= (\dot{X}'\dot{X})^{-1}\dot{X}'\dot{y} \\ \hat{\beta}_{BW} &= (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{y}\end{aligned}$$

3. What happens to the relative weights on the "within" and "between" estimators as we increase the sample size, i.e. $T \to \infty$?

   As $T \to \infty$, $\lambda \to 0$. This implies:

$$\hat{\beta}_{GLS} \to [\dot{X}'\dot{X}]^{-1}(\dot{X}'\dot{y}) = \hat{\beta}_{FE}$$

   So, so full weight is given to fixed effects estimator, and $\hat{\beta}_{GLS}$ behaves like $\hat{\beta}_{FE}$

4. Suppose that the random effects assumption $\mathbb{E}\big[\alpha_i|x_{i1,\dots,x_{iT}}\big] = 0$ does not hold. Characterize the bias of the estimators $\hat{\beta}_{FE}$, $\hat{\beta}_{W}$. (Note: An estimator $\hat{\beta}$ is unbiased if $\mathbb{E}\big[\hat{\beta}\big] = \beta$)

$$\mathbb{E}\big[\hat{\beta}_{FE}|X\big] = \mathbb{E}\big[(\dot{X}'\dot{X})^{-1}\dot{X}'\dot{y}|X\big]$$

   where, $\dot{y} = \dot{X}\beta + \epsilon$.

$$\begin{aligned}\mathbb{E}\big[\hat{\beta}_{FE}|X\big] &= \mathbb{E}\big[(\dot{X}'\dot{X})^{-1}\dot{X}'(\dot{X}\beta + \epsilon)|X\big] \\ &= \mathbb{E}\big[(\dot{X}'\dot{X})^{-1}\dot{X}'\dot{X}\beta|X\big] + \mathbb{E}\big[(\dot{X}'\dot{X})^{-1}\dot{X}'\epsilon)|X\big] \\ &= \beta + (\dot{X}'\dot{X})^{-1}\dot{X}'\mathbb{E}\big[\epsilon|X\big] \\ &= \beta\end{aligned}$$

   since $\mathbb{E}\big[\epsilon|X\big] = 0$. Therefore, fixed effects estimator is unbiased.

$$\mathbb{E}\big[\hat{\beta}_{BW}|X\big] = \mathbb{E}\big[(\bar{X}'\bar{X})^{-1}\bar{X}'\bar{y}|X\big]$$

   where, $\bar{y} = \bar{X}\beta + \alpha + \epsilon$.

$$\begin{aligned}\mathbb{E}\big[\hat{\beta}_{BW}|X\big] &= \mathbb{E}\big[(\bar{X}'\bar{X})^{-1}\bar{X}'(\bar{X}\beta + \epsilon + \alpha)|X\big] \\ &= \mathbb{E}\big[(\bar{X}'\bar{X})^{-1}\bar{X}'\bar{X}\beta|X\big] + \mathbb{E}\big[(\bar{X}'\bar{X})^{-1}\bar{X}'\epsilon + \alpha)|X\big] \\ &= \beta + (\bar{X}'\bar{X})^{-1}\bar{X}'\mathbb{E}\big[\epsilon + \alpha|X\big] \\ &= \beta + (\bar{X}'\bar{X})^{-1}\bar{X}'\mathbb{E}\big[\alpha|X\big]\end{aligned}$$

   Note that the between estimator will be biased, since $\mathbb{E}\big[\alpha|X\big] \neq 0$.

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 $T \to \infty$.

$$\hat{\beta}_{GLS} = R\hat{\beta}_{FE} + (I - R)\hat{\beta}_{BW}$$

$$\begin{aligned}
\mathbb{E}\big[\hat{\beta}_{GLS}|X\big] &= R\mathbb{E}\big[\hat{\beta}_{FE}|X\big] + (I-R)E\hat{\beta}_{BW}|X \\
&= R\beta + (I-R)(\beta + (\bar{X}'\bar{X})^{-1}\bar{X}'\mathbb{E}\big[\alpha|X\big])
\end{aligned}$$

Since $\hat{\beta}_{GLS} \to \hat{\beta}_{FE}$ as $T \to \infty$, bias will tend to zero (asymptotically consistent).

Problem 2 (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?

The Likelihood function is:

$$\begin{aligned}
L(\gamma|x) &= \prod_{i=1}^{N} f(x_i) \\
&= \prod_{i=1}^{N} \gamma x^{\gamma-1}\exp(-x^{\gamma}) \\
&= \gamma^N \exp(-\sum_{i=1}^{N} x^{\gamma}) \prod_{i=1}^{N} x^{\gamma-1}
\end{aligned}$$

Taking the log, we get the log-likelihood function:

$$l(\gamma|x) = N ln(\gamma) - \sum_{i=1}^{N} x^{\gamma} + (\gamma - 1)\sum_{i=1}^{N} ln(x)$$

2. Calculate the gradient (or first derivative) of your log-likelihood function.

$$\frac{\partial l}{\partial \gamma} = \frac{N}{\gamma} - \sum_{i=1}^{N} ln(x_i)x_i^{\gamma} + \sum_{i=1}^{N} ln(x_i)$$

3. Using the first order condition, what is the MLE estimator for γ?

$$\frac{\partial l}{\partial \gamma} = \frac{N}{\gamma} - \sum_{i=1}^{N} ln(x_i)x_i^{\gamma} + \sum_{i=1}^{N} ln(x_i) = 0$$

There doesn't seem to be a closed form solution for γ

4. Verify that the second order condition guarantees a unique global solution.

$$\frac{\partial^2 l}{\partial \gamma^2} = -\frac{N}{\gamma^2} - \sum_{i=1}^{N} (ln(x_i))^2 x_i^{\gamma} < 0$$

The second derivative is always negative, which guarentees an unique global maximum.

5. 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.

```
mle_weibull = function(X, g){
  N = length(X)
  l = N*log(g) - sum(X^g) + (g-1)*(sum(log(X)))

  # return negative, since optim function minimizes instead of maximizes.
  return(-l)

}
```

6. 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.

```
load('data/prob_4_simulation.rda')
optim(par=1, fn=mle_weibull, X=sim)
```

```
## Warning in optim(par = 1, fn = mle_weibull, X = sim): one-dimensional optimization by Nelder-Mead is
## use "Brent" or optimize() directly

## $par
## [1] 1.982227
##
## $value
## [1] 603.5963
##
## $counts
## function gradient
##       30       NA
##
## $convergence
## [1] 0
```

```
##
## $message
## NULL
```

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.

```r
gradient = function(X, g){
  N = length(X)
  d = N/g - sum(log(X)*(X^g)) + sum(log(X))
  return(d)
}
```

```r
optim(par=1, fn=mle_weibull, gr=gradient, X=sim)
```

```
## Warning in optim(par = 1, fn = mle_weibull, gr = gradient, X = sim): one-dimensional optimization by
## use "Brent" or optimize() directly

## $par
## [1] 1.982227
##
## $value
## [1] 603.5963
##
## $counts
## function gradient
##       30       NA
##
## $convergence
## [1] 0
##
## $message
## NULL
```

c. Compare both the number of iterations until convergence and your estimated γ values from both runs.

optim doesn't seem to be using the gradient function. The number of iterations are same in both cases (30).

However, the optimal value $\gamma_{\hat{MLE}} = 1.982227$ seems to be correct:

```r
print(gradient(sim, 1.5))
```

```
## [1] 262.2413
```

```r
print(gradient(sim, 1.982227))
```

```
## [1] 0.03797449
```

```r
print(gradient(sim, 2.5))
```

```
## [1] -230.3122
```