ECMA 31130 HW2

Yijing Zhang & Jeffrey Wang

Question 1

Because from the specification of $\log R_i$ that ξ_i represents all the unobserved variables that determine the individual i's non-labour income, the parameter a of the model captures or leaves out the effect of these unobserved characteristics on the individual's disutility of labor.

```
p  = list(gamma = 0.8,beta=1,a=1,rho=1,eta=0.2,delta=-0.2,delta0=-0.1,nu=0.5) # parameters
N=10000  # size of the simulation
set.seed(123456)

simdata = data.table(i=1:N,X=rnorm(N))

# simulating variables
simdata[,X := rnorm(N)]
simdata[,Z := rnorm(N)]
simdata[,u := rnorm(N)]
simdata[,lw := p$eta*X  + Z + 0.2*u ]  # log wage

simdata[,xi := rnorm(N)*0.2]
simdata[,lr := lw + p$delta0+ p$delta*Z + xi]; # log home productivity

simdata[,eps:=rnorm(N)*0.2]
simdata[,beta := exp(p$nu*X  + p$a*xi + eps)]; # heterogenous beta coefficient

# compute decision variables
simdata[, lfp := log(p$rho) + lw >= lr] # labor force participation
simdata[, h   := (p$rho * exp(lw)/beta)^(1/p$gamma)] # hours
simdata[lfp==FALSE,h:=NA][lfp==FALSE,lw:=NA]
simdata[,mean(lfp)]

simdata[, lh := log(h)]
```

Question 2

For a = 0:

```
pander(summary(lm(lh ~ lw + X, simdata)))
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.001408	0.003429	-0.4106	0.6814
lw	1.246	0.003594	346.7	0
X	-0.6268	0.003202	-195.8	0

Observations	Residual Std. Error	R^2	Adjusted R^2
6408	0.2487	0.9545	0.9545

For a = 1:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.1415	0.004356	32.49	1.851e-214
lw	1.15	0.004565	251.9	0
X	-0.6111	0.004068	-150.2	0

Observations	Residual Std. Error	R^2	Adjusted R^2
6408	0.316	0.9187	0.9187

We observe that the coefficients for log wage and X have larger absolute values when we set parameter $a = 0$ than when we set $a = 1$. Because if $a \neq 0$, then the term $a\xi_i$ becomes an omitted variable that enters the regression of log hours on log wage and X, which will absorb its effect on the individual's labour decision from the coefficients of log wage and X. In other words, if an individual's disutility of labour becomes correlated with her non-labour income, then the coefficients in front of log wage and X are biased, since we no longer have the exogeneity condition on ϵ_i.

Question 3

We begin by deriving the FOC and its log linear form when $e = 1$.

Call $\tilde{h}^*$ and $\tilde{w}^*$ the log observed observed hours and consumption when $lfp = 1$, we write out our log linear form and replace β with our specification:

$$\tilde{h}^* = \frac{1}{\gamma}\log\rho + \frac{1}{\gamma}\tilde{w}^* - \frac{1}{\gamma}(\nu x_i + \epsilon_i + a\xi_i)$$

$$\tilde{h}^* = \frac{1}{\gamma}\tilde{w}^* - \frac{\nu}{\gamma}x_i - \frac{\epsilon_i}{\gamma} - \frac{a\xi_i}{\gamma} + \frac{1}{\gamma}\log\rho$$

Now we take the conditional expectation of hours based on everything we observe

$$E[\tilde{h}^*|\tilde{w}, x_i, z_i, lfp = 1] = \frac{1}{\gamma}\tilde{w} - \frac{\nu}{\gamma}x_i - \frac{1}{\gamma}E[\epsilon_i + a\xi_i|\tilde{w}, x_i, z_i, lfp = 1] + \frac{1}{\gamma}\log\rho$$

We now take a better look at our heckman correction term. Since ϵ_i is independent and mean zero, ϵ_i is independent and mean zero,

$$E[\epsilon_i|\tilde{w}, x_i, z_i, lfp = 1] = 0$$

$$E[a\xi_i|lfp = 1] = E[\xi_i|\log w_i + \log\rho > \log R] = E[\xi_i|\log\rho > \delta_0 + \delta z_i + \xi_i] = aE[\xi_i|\xi_i < \log\rho - \delta_0 - \delta z_i]$$

Now let's take $\rho = 1$ to make $\log\rho = 0$, we express the inverse mills ratio:

$$\lambda_i = E[\xi_i|\xi_i < -\delta_0 - \delta z_i] = -\frac{\sigma_{\xi_i}\Phi'(\frac{-\delta_0 - \delta z_i}{\sigma_{\xi_i}})}{\Phi(\frac{-\delta_0 - \delta z_i}{\sigma_{\xi_i}})}$$

We do a probit of lfp on z_i to get δ_0 and δ.

```
fit2 = glm(lfp ~ Z,simdata,family = binomial(link = "probit"))
pander(summary(fit2))
```

Question 4

#construct the inverse Mills ratio.

```
simdata[, ai := predict(fit2)]
simdata[, m := dnorm(ai)/pnorm(ai)]

fit3 = summary(lm(lh ~ lw + X + m,simdata))
pander(fit3)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.02577	0.01658	-1.554	0.1202
lw	1.267	0.01117	113.4	0
X	-0.6266	0.004666	-134.3	0
m	0.293	0.02836	10.33	8.009e-25

Table 6: Fitting linear model: lh ~ lw + X + m

Observations	Residual Std. Error	R^2	Adjusted R^2
6405	0.3167	0.921	0.921

We see that our estimated coefficients indeed correspond to our results before when we set $a = 0$. But now our model have $a = 1$.

Question 5

From Q3 we have

$$E[\xi_i|\xi_i < -\delta_0 - \delta z_i]$$

And since we have $\xi_i \sim -exp$, we can write

$$E[\xi_i|\xi_i < -\delta_0 - \delta z_i] = 1 - \delta_0 - \delta z_i$$