# 1 Heckman selection model

1. You want to model the wage equation for women

   You consider estimating the model:

$$\ln wage = \alpha + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \varepsilon \tag{1}$$

   Read the data into Stata:

```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/mroz.dta

. label var nwifeinc "Income not from wife"

. label var kidslt6 "Kids 1-6 years old"

. label var kidsge6 "Kids > 6 years old"
```

   (a) Estimate the model in equation 1 using OLS. Assuming that the RHS-variables are all uncorrelated with the error term in the population, why may the estimates still be biased?

   (b) Estimate the Heckman selection model in two individual steps without exclusion restrictions, by predicting the inverse mills ratios and including this as a control variable in the wage equation.

   (c) Estimate the Heckman selection model using the command -heckman-, and include the variables -nwifeinc-, -age-, -kidslt6-, and -kidsge6- in the selection model. Do you think the exclusion restriction is plausible?

   (d) Reestimate the model in 1c using the twostep estimator (-heckman, twostep-) and in individual steps (as in 1b above), and compare the estimates.

# 2 Multinomial Choice

1. Now you know want to model whether people go fishing from the beach, pier, private boat, or charter boat.

   You consider estimating the *conditional Logit model*:

$$\Pr(y_i = j \mid x_{ij}) = \frac{\exp(x'_{ij}\beta)}{\sum_{k=1}^{K}\exp(x'_{ik}\beta)}$$

   the *multinomial Logit model*

$$\Pr(y_i = j \mid w_i) = \frac{\exp(w'_i\gamma_j)}{\sum_{k=1}^{K}\exp(w'_i\gamma_k)}$$

   and the *mixed Logit model*

$$\Pr(y_i = j \mid x_{ij}, w_i) = \frac{\exp(x'_{ij}\beta + w'_i\gamma_j)}{\sum_{k=1}^{K}\exp(x'_{ik}\beta + w'_i\gamma_k)}$$

   Again, read the data into Stata:

```
. use http://fmwww.bc.edu/ec-p/data/mus/mus15data
```

   As potential regressors you have income (varies across individuals but not across alternatives), and price and catch rate (varies across individuals and alternatives).

   (a) Calculate some relevant descriptive statistics (think about what you would like to know).

   (b) Estimate the multinomial Logit model (using -mlogit-). What are the sample partial effects of your regressor(s)? What are these partial effects for those who chose different modes of fishing?

The IIA property of the Multinomial Logit model implies that

$$\Pr(y_i = j \,|\, y_i = k \text{ or } y_i = j,\, w_i) = \frac{p_{ij}}{p_{ij} + p_{ik}} = \frac{\exp(w_i'(\beta_j - \beta_k))}{1 + \exp(w_i'(\beta_j - \beta_k))} \tag{2}$$

which is a Logit model.

(c) Estimate (2) for different $j$'s and compare your estimates to (b). What do you conclude? (Think about your choice of $k$).

(d) Estimate the conditional Logit model (using -clogit-). What are the sample partial effects of your regressor(s)? What are these partial effects for those who chose different modes of fishing?

Assume now that fishing consists of three *ordered* alternatives; $y_i = 0$ is fishing from a pier or beach, $y_i = 1$ is fishing from a private boat, and $y_i = 2$ is fishing from a charter boat.

(e) Estimate an ordered Logit model (using -ologit-) with price as the only regressor, and provide an interpretation of the estimated coefficient.