Data analysis chapter example

This chapter will focus on the results of the data analysis. The first section will discuss the descriptive statistics and in the second section the results of the Heckman two-step approach will be discussed.

Descriptive statistics

The descriptive statistics of the survey data will be discussed by comparing and characterizing the households that affected and do not affected by the climate change. The sample size used for the analysis is therefore 420 respondents.

An uneven distribution of land ownership exists in coastal region of Bangladesh, with a significant proportion of land being owned by large landowners (Alauddin and Hamid 1997). Agricultural survey (1996) shows that 54% of families in coastal areas hold only 17% of the total agricultural land (PDO-ICZMP-2003). The majority of the rural population is either landless farmers (who sell their labor or cultivate other's land)or marginal farmers (who have less than . 2 ha of property) (Opstal 2006). Over the past decade the farmers are declined. Now a day in the coastal Bangladesh fishing is one of the most important economic activities. They are mostly landless or have a small plot of land to use for living purpose.

In the study area total land size is changed due to climate change. It appears from the given table which shows the comparative analysis of land pattern before and after Aila. In 2008 the average amount is 157. 02 hectare/year and in 2009 it is 99. 89 hectare/year. Land is used for different purposes. In 2008, 159 respondents used their land for cultivation i. e they are the

agricultural land owner and due to climate change only 75 respondents are the owner of the agricultural land. This amount is decreasing.

In last 5 years 62 households lost their land in the study area. The total amount of damaged land is 36911. 58 hectares. Most of the people depend on agriculture so this is a great loss for their survive. For this their income is decreased, expenditure is decreased and they have no enough money to buy the agricultural land. From this it is concluded that they live below poverty line. According to a recent (Oct'09) study done by the South Asia Association of Poverty Eradication, each affected household has seen their income decrease by approximately 44% as a result of Cyclone Aila.

The main independent variable is expenditures by household for a basket of basic needs, which is considered as a measurement of ' poverty.' This expenditure measurement actually represents a poverty threshold value, which is derived from HIES (Household Income-Expenditure Survey 2009) by BBS and is equivalent to US$ 208/capita/year (BBS, 2008). It is referred as ' Basic Need Cost' in the model.

In 2009 we get only 84 respondents out of 420 do not live below poverty line. It is estimated by using our expenditure data from primary survey analysis. So due to climate change most of the households live below poverty line.

Econometric Analysis

Now we would like to continue with figuring out the nature and extent of relationship between agricultural land ownership pattern and poverty of Koyra. Hence, in this chapter we conduct econometric analysis.

Variables used in econometric models

With a view to identifying the relationship pattern between agricultural land ownership pattern and poverty we ran a number of econometric models. But before we proceed to the operation with econometric models, let us have a look at the variables used in the model.

Dependent variable

The dependent variable is total land owned by, which is considered to be affected by climate change. This variable indicates how much land was owned by the household in 2009. The values were taken in hectares for the entire household.

Independent variables

Below we have mentioned the independent variables, with short explanation, that we used in models. Variable ' household size' refers to the total number of members in a household. ' Education' refers to household's average aggregate academic schooling year. It is the number obtained by summing up of formal schooling years of all members in a household and then dividing it with the number of total household members. This variable is considered as a proxy for capacity of households. The variable ' Duration with community' refers to the number of years the respondent household living with the current community.

Along with the above-mentioned dependent and independent variables, we used the following two independent variables for constructing correlation and regression.

Econometric Methodology:

We used a Heckman Two Step Model for dependent variable ' land ownership' in order to find out if there is any sample selection bias in the model. This model consists of two processes that are addressed by two different equations: a selection equation and a conditional equation. The first probit equation is a selection process for the households having land-ownership or not. In the second equation the effects of independent variables on ' land ownership' are examined.

These processes are related to each other through their error terms which contain the unobservable. If there is no correlation between the error terms of the two equations, there is no need to perform a Heckman two step approach as there is no sample selection bias and an OLS regression provides the unbiased result (Dow and Norton, 2003).

The Heckman two-step approach is based on the assumption that the selection equation and the conditional equation are related to each other through their error terms. When there is no relation between the error terms there is no need to perform a Heckman two step approach as there is no sample selection bias and an OLS regression will give unbiased estimators. For such a model, the bottom line in STATA output gives a value for ρ (rho) with associated p-value. This ρ is a likelihood ratio indicating the correlation between the error terms of the equations in Heckman model.

The correlation between the error terms is indicated in table (Annex) by the selectivity parameter, ρ. The Heckman's lambda is included in the regression to control for the influence of unobserved characteristics of the variables.

The regression coefficient of the control factor is an indicator for the covariance of the error terms. In the model the control factor is non-significant.

The missing data problem can arise in a variety of forms. We can see that there are missing data in the sample. The number of missing data in is 3, but the problem is more severe for, where the number of missing data is 80. Since the data is missing mainly on the dependent variable, a nonrandom sample selection exists in this case. There is a possibility that due to some common pattern, the respondents did not provide any data. If that has happened, bias could always occur in OLS in estimating the population model. As a result, we use here the Heckman model.

Our model is

Empirical results

This chapter will focus on the results of the data analysis. The first section will discuss the descriptive statistics and in the second section the results of the Heckman two-step approach will be discussed.

Descriptive statistics

The descriptive statistics of the survey data will be discussed by comparing and characterizing the households that affected and do not affected by the climate change. The sample size used for the analysis is therefore 420 respondents.

An uneven distribution of land ownership exists in coastal region of Bangladesh, with a significant proportion of land being owned by large

landowners (Alauddin and Hamid 1997). Agricultural survey (1996) shows that 54% of families in coastal areas hold only 17% of the total agricultural land (PDO-ICZMP-2003). The majority of the rural population is either landless farmers (who sell their labor or cultivate other's land)or marginal farmers (who have less than . 2 ha of property) (Opstal 2006). Over the past decade the farmers are declined. Now a day in the coastal Bangladesh fishing is one of the most important economic activities. They are mostly landless or have a small plot of land to use for living purpose.

In the study area total land size is changed due to climate change. It appears from the given table which shows the comparative analysis of land pattern before and after Aila. In 2008 the average amount is 157. 02 hectare/year and in 2009 it is 99. 89 hectare/year. Land is used for different purposes. In 2008, 159 respondents used their land for cultivation i. e they are the agricultural land owner and due to climate change only 75 respondents are the owner of the agricultural land. This amount is decreasing.

In last 5 years 62 households lost their land in the study area. The total amount of damaged land is 36911. 58 hectares. Most of the people depend on agriculture so this is a great loss for their survive. For this their income is decreased, expenditure is decreased and they have no enough money to buy the agricultural land. From this it is concluded that they live below poverty line. According to a recent (Oct'09) study done by the South Asia Association of Poverty Eradication, each affected household has seen their income decrease by approximately 44% as a result of Cyclone Aila.

The main independent variable is expenditures by household for a basket of basic needs, which is considered as a measurement of ' poverty.' This expenditure measurement actually represents a poverty threshold value, which is derived from HIES (Household Income-Expenditure Survey 2009) by BBS and is equivalent to US$ 208/capita/year (BBS, 2008). It is referred as ' Basic Need Cost' in the model.

In 2009 we get only 84 respondents out of 420 do not live below poverty line. It is estimated by using our expenditure data from primary survey analysis. So due to climate change most of the households live below poverty line.

Econometric Analysis

Now we would like to continue with figuring out the nature and extent of relationship between agricultural land ownership pattern and poverty of Koyra. Hence, in this chapter we conduct econometric analysis.

Variables used in econometric models

With a view to identifying the relationship pattern between agricultural land ownership pattern and poverty we ran a number of econometric models. But before we proceed to the operation with econometric models, let us have a look at the variables used in the model.

Dependent variable

The dependent variable is total land owned by, which is considered to be affected by climate change. This variable indicates how much land was owned by the household in 2009. The values were taken in hectares for the entire household.

Independent variables

Below we have mentioned the independent variables, with short explanation, that we used in models. Variable ' household size' refers to the total number of members in a household. ' Education' refers to household's average aggregate academic schooling year. It is the number obtained by summing up of formal schooling years of all members in a household and then dividing it with the number of total household members. This variable is considered as a proxy for capacity of households. The variable ' Duration with community' refers to the number of years the respondent household living with the current community.

Along with the above-mentioned dependent and independent variables, we used the following two independent variables for constructing correlation and regression.

Econometric Methodology:

We used a Heckman Two Step Model for dependent variable ' land ownership' in order to find out if there is any sample selection bias in the model. This model consists of two processes that are addressed by two different equations: a selection equation and a conditional equation. The first probit equation is a selection process for the households having land-ownership or not. In the second equation the effects of independent variables on ' land ownership' are examined.

These processes are related to each other through their error terms which contain the unobservable. If there is no correlation between the error terms of the two equations, there is no need to perform a Heckman two step

approach as there is no sample selection bias and an OLS regression provides the unbiased result (Dow and Norton, 2003).

The Heckman two-step approach is based on the assumption that the selection equation and the conditional equation are related to each other through their error terms. When there is no relation between the error terms there is no need to perform a Heckman two step approach as there is no sample selection bias and an OLS regression will give unbiased estimators. For such a model, the bottom line in STATA output gives a value for ρ (rho) with associated p-value. This ρ is a likelihood ratio indicating the correlation between the error terms of the equations in Heckman model.

The correlation between the error terms is indicated in table (Annex) by the selectivity parameter, ρ. The Heckman's lambda is included in the regression to control for the influence of unobserved characteristics of the variables. The regression coefficient of the control factor is an indicator for the covariance of the error terms. In the model the control factor is non-significant.

The missing data problem can arise in a variety of forms. We can see that there are missing data in the sample. The number of missing data in is 3, but the problem is more severe for , where the number of missing data is 80. Since the data is missing mainly on the dependent variable, a nonrandom sample selection exists in this case. There is a possibility that due to some common pattern, the respondents did not provide any data. If that has happened, bias could always occur in OLS in estimating the population model. As a result, we use here the Heckman model.

Our model is

We assumed that is observed if

Where and have correlation

Results:

The results of our Heckman model are provided in Table (Annex). Using as a dependent variable in Heckman regression, we find and the constant term are significant while is insignificant. We also find positive relationship for and with . Considering the absolute values of the coefficients (table), the result shows that is the most influential between the two variables.

A typical use of a logarithmic transformation variable is to pull outlying data from a positively skewed distribution closer to the bulk of the data in a quest to have the variable be normally distributed. In regression analysis the logs of variables are routinely taken, not necessarily for achieving a normal distribution of the predictors and/or the dependent variable but for interpretability.

The standard interpretation of coefficients in a regression analysis is that a one unit change in the independent variable results in the respective regression coefficient change in the expected value of the dependent variable while all the predictors are held constant. Interpreting a log transformed variable can be done in such a manner; however, such coefficients are routinely interpreted in terms of percent change (Introductory Econometrics: A Modern Approach by Woolridge for discussion and derivation).

We'll explore the relationship between the landownership pattern and the per capita consumption expenditure. In this model we are going to have the dependent variable in its original metric and the independent variable log-transformed. Similar to the prior example the interpretation has a nice format, a one percent increase in the independent variable increases (or decreases) the dependent variable by (coefficient/100) units. In this particular model we take log with PCE and the coefficients on and represent the estimated marginal effects of the regressors in the underlying regression equation. So, an increase in the household size by one member increases land ownership by 6. 30 hectares and an increase in the household consumption expenditure by one percent increases land ownership by 0. 613 hectares.

On the other hand, household size is the least influential variable. It is positively related with landownership pattern. So these two variables have greater influence on poverty. We used the Heckman two step models while taking land ownership as a dependent variable in the conditional equation of this model, along with other independent variables, result in model shows that PCE is positively related with landownership.

The p value of lambda is 0. 193 i. e. 19%. So this is not significant for the model i. e. there is no correlation between the error terms of the two equations in Heckman model. The lambda term is positively signed – which suggests that the error terms in the selection and primary equations are positively correlated. So (unobserved) factors that make more observable tend to be associated with higher values of our independent variables in the

selection equation. However, since the lambda term is not significant, we cannot come to any such conclusion and hence we conducted OLS.

But if we use the OLS we get the following

Table 1: OLS Result

_____

___

lnd_owners~p | Coef. Std. Err. t P>| t| [95% Conf. Interval]

————————-+
_____-

lnpce | 58. 21023 18. 98437 3. 07 0. 002 20. 86622 95. 55423

hh_size | 4. 660069 6. 495749 0. 72 0. 474 -8. 117666 17. 4378

_cons | -204. 742 97. 52465 -2. 10 0. 037 -396. 5819 -12. 90203

_____

___

We present the usual OLS regression in Table 1. As we can see from Table 1, and is both positive, while the former is not significant and the latter is significant. Similarly, the constant term is negative but significant.

Table 2

From the above OLS table we consider the independent variables are per capita expenditure, education level, during with the community, household size and asset 2008 and the dependent variable is land ownership pattern of the respondents. In this analysis the model is significant in case of asset

2008 for dependent variable land ownership because in this case the value of P is 0%. We know if the value of P is less than 5% then the model is significant. From the regression we get per capita expenditure, education level, during with the community and asset 2008 is positive. But without asset 2008 all other variables are not significant. Similarly the constant term is also positive but not significant.

Results from various OLS regression models are shown in Table 1 and. 2. The former shows results when model is run with and while the latter shows results when land ownership is incorporated with other independent variables. Values of coefficient are different for the independent variables in the result tables. Using land ownership (i. e. our measure of poverty) as a dependent variable in OLS regression, we found without one, all the explanatory variables are not significant (Table 2). We also found significant positive relationship per capita expenditure, education level, during with the community and asset 2008 with land ownership whereas it is significantly negative for household size.

Annex

. heckman lnd_ownership lnpce hh_size, twostep select(lnpce edulevel duringwithcomty hh_size asst2008) rhosigma

Heckman selection model — two-step estimates Number of obs = 417

(regression model with sample selection) Censored obs = 80

Uncensored obs = 337

Wald chi2(4) = 9. 83

Prob > chi2 = 0. 0434

_____

___

| Coef. Std. Err. z P>| z| [95% Conf. Interval]

————————-+ _____-

lnd_owners~p |

lnpce | 61. 28878 20. 67387 2. 96 0. 003 20. 76873 101. 8088

hh_size | 6. 303549 7. 203314 0. 88 0. 382 -7. 814687 20. 42179

_cons | -286. 9731 123. 3481 -2. 33 0. 020 -528. 731 -45. 21517

————————-+ _____-

select |

lnpce | . 0682579 . 1348031 0. 51 0. 613 -. 1959514 . 3324671

edulevel | . 0096151 . 025462 0. 38 0. 706 -. 0402896 . 0595197

duringwith~y | . 0161874 . 005286 3. 06 0. 002 . 005827 . 0265477

hh_size | . 007615 . 046654 0. 16 0. 870 -. 0838252 . 0990552

asst2008 | -1. 13e-06 7. 34e-07 -1. 53 0. 125 -2. 57e-06 3. 12e-07

_cons | -. 0686488 . 6543009 -0. 10 0. 916 -1. 351055 1. 213757

—————-+

_____-

mills |

lambda | 181. 4302 139. 4798 1. 30 0. 193 -91. 94525 454. 8057

—————-+

_____-

rho | 0. 74328

sigma | 244. 09453

lambda | 181. 43021 139. 4798

_____

__