CHAPTER 3

Methodology

3.1 Conceptualize of the study

This research was conducted in order to analyze data from launching online questionnaires and samplings to specific groups of informants in Cambodia. The aims of this study are to investigate the role of limited rationality on the decision making with the Thais, and possibly would explore the details of information between Cambodia and Thailand. The procedures of the investigation: at first, there is a pre-existing perception of Cambodians toward the Thais, afterwards the information is being examined by spreading questionnaires as qualitative analysis and those questionnaires were taken to analyze by employing two main econometric models to meet with two main objectives stated in the introductory part such as Logit and Ordered Probit Models. It is to evaluate Cambodia decision makers in terms of establishing friendship and attempting to get involved with the Thais, and characteristics analysis under rationality theory. The organizational framework of the study as followings:

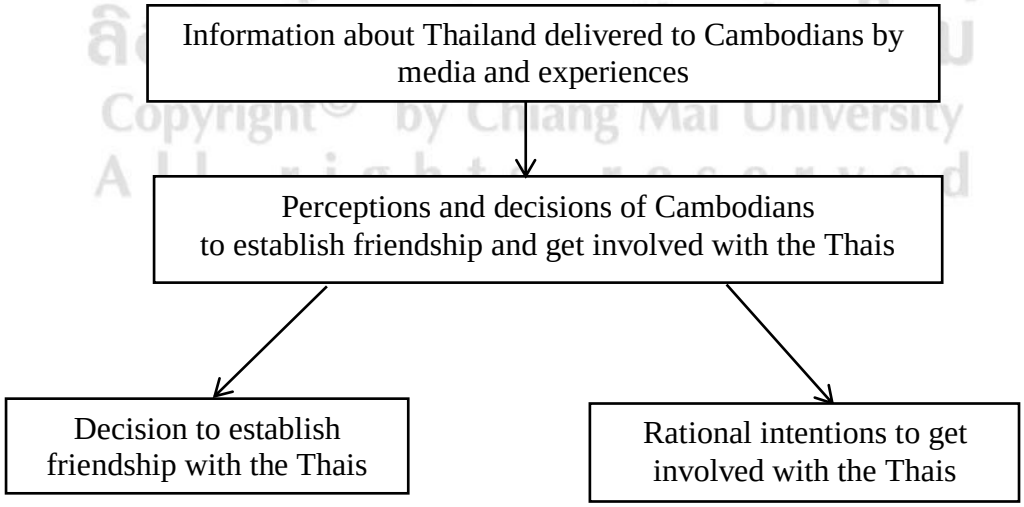

Figure 1.1 Conceptual Framework of the study

3.2 Research Methodologies

This research was conducted in the primary level in Cambodia, in which respondents have accessed to the internet, specifically to Cambodian citizens. Online questionnaires were launched electronically from the initial date of approved proposal and questionnaire from the faculty committee. The methodology flows in conducting the research as following:

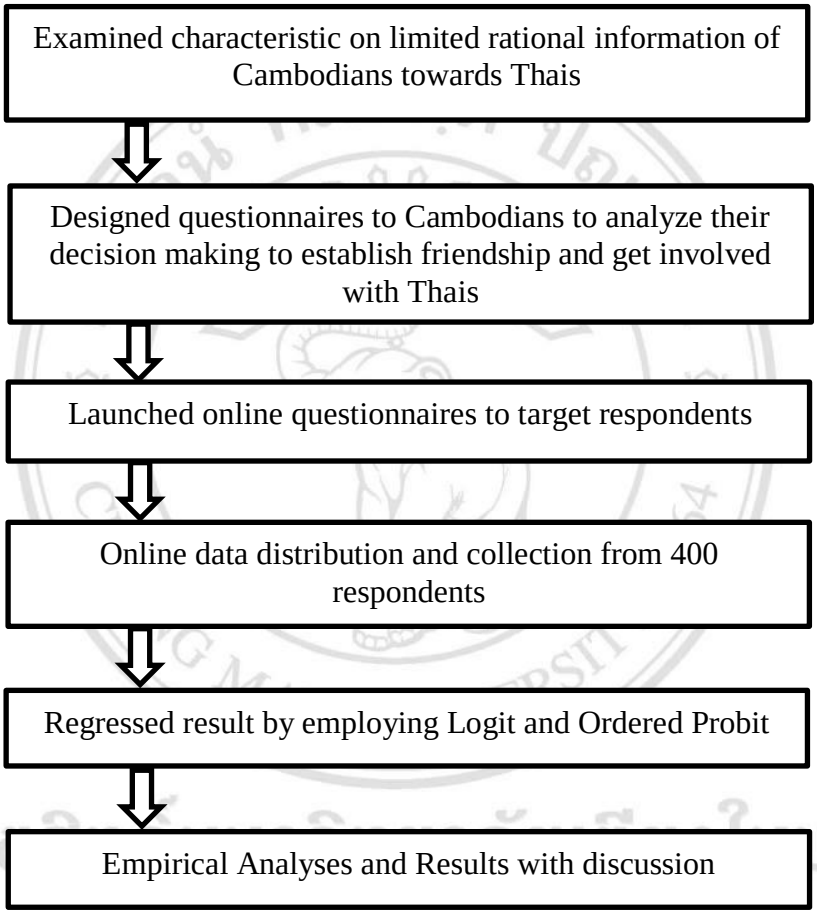

Figure 3.2 Research Flow

Figure 3.2 described precisely on the flows of conducting research survey for this study. Questionnaire is designed to test on Cambodians existing perceptions toward Thailand as well as Thai people, then being examined their characteristics. With advancement of technology helps ease in conducting this research through online efficiently. First of all, questionnaire designed to meet the objectives of the study on the current perceptions between Cambodia and Thailand socially and economically. Questionnaires distributed to the sample of population as target respondents (Cambodia population). According to

Yamane's formula, (1967) sample size can be determined by 400 samples. Data collection was conducted by online survey website. Afterwards, data was tested through two econometric models in order to regress empirical studies and analyzations by employing Logit and Ordered Probit Models based on the two objectives in this research study.

3.2.1 Using Yamane Formula Calculating a Sample Size

Population surveyed in this research was Cambodians online users, tactically in Cambodia nationwide. A report from internet world statistic in 2013 mentioned there were 927, 500 internet users in Cambodia while Facebook subscribers were 732, 220, though the figures keep increasing every day, this amount is be taken as population size in this study. It is big enough to resemble the entire population since the number of users or subscribers are dispersed over the country. Technically, there are lots formulas to adopt for calculating the sample sizes in the survey. However, this research has adopted a simple formula from Yamane's formula (1967), which postulated the simplified formula proportion to calculate the sample sizes of proportions as below equation. This formula is the best to determine the minimal sample size for the given population size. By using Yamane's formula (1967) of the sample size of population, it is commonly assumed that the degrees of variability proportion is 0.5 (P=0.5) with a confidence level of coefficient 95% and with an error of 5%.

$$n = \frac{\mathrm{N}}{1 + N(e)^2} \qquad (1)$$

Where, n is sample size, N is population size, e is the level of precision which is acceptable sampling errors, where errors of 5 percentage points, then we plug the number in the equation 4, we get:

$$n = \frac{927,\ 500}{1 + 927,\ 500 * (0.05)^2} = 399.82 \qquad (2)$$

Based on the formula, the calculation of wanted respondents are from the 400 sample sizes, which are rounded up from the figure above of the entire population with cyber user respondents to the questionnaires distributed. Response rate is the survey is considered as the most crucial part in questionnaires survey. By definition, the percentage of people who did actually fill out a survey that they receive is known as response rate.

Consequently, response rate has to be taken seriously into account to get good estimation and determination of the total numbers of the survey in order to obtain the numbers of wanted respondents in conducting the survey and get them as the sample sizes in this study. Financially, it is always recognized that to get the complete and perfect information is hardly possible and always costly. As a consequence, this research was conducted as the survey to reach out the respondents directly in order to get to analyze sample size of Cambodians' attitude toward Thailand with more efficiency. For online survey, it is widely known that the response rate between 20 to 30 percent is considered as highly successful response rate in launching online survey. Questionnaire distribution provided the respondent's rates higher than 60 percent over all since the conducted questionnaires were mainly for the active online users on social networking, which could be considered as a successful data collection from the survey.

3.2.2 Data Source and Data Description

Questionnaires were electronically distributed through online database users in order to reach out target respondents in Cambodia. Questionnaires were designed for the survey to generalize results from the sample of Cambodian population, within 1 period by the end of May, 2015. Initially, questionnaires were launched and distributed to Cambodia's online population. This survey was conducted via social media network to target samples of active online users. In order to ensure sample size was enough and appropriate in using and running in the regression, the distribution was more than target volumes as the samples. The result from collection was 415 samples, which were the suitable amount of samples based on the Yamane (1967). Estimated around 14 percent of respondents skipped the questions, which the entire observations dropped from model to model regressed. Apart from that, the investigation on respondents' decision-making through the information they provided with 5-point likert scale analysis, whether or not their responses were under bounded rationality theory based on the choice they made. It is important in transferring and analyzing Cambodians' perception from the initial survey to final results in the regression by employing econometric models under economics theory of bounded rationality.

The questionnaires originally used in English. Data description from the survey was divided into 3 sections. Frist section was described with demographic variables such

as gender, marital status, age, education, occupation, income. Second section part A inquired respondents' travel experience and perception toward Thailand, length of their stay while traveling in Thailand, their purposes of traveling, and whether or not they have Thai friends from their trip. This section was optional for those never had traveling experience to Thailand at all, which meant they could skip this part and jump to the next section. Additionally, Part B, there were a list of factors to ask them about terms of their decision making with Thais, which consisted of thirteen sub statements. Due to all statements they decided, the last decision for them to answer was whether or not they would like to establish friendship with the Thais. The final section of the questionnaire was about intention statements contained 9 sub statements to get involved with Thais. The mode of the statements enabled respondents to rate their preference by using 5-point Likert scales from strongly disagree scored as 1, disagree scored as 2, neutral scored 3, agree scored 4, and strongly agree scored as 5.

3.3 Models Specifications

Econometric methods had been regressed according to the two set based of main objectives of this research in order to give better empirical analyzing and unbiased testing. The predictions of models were being tested by providing the empirical evidence on that. This study employed econometric investigation on the effects of Cambodians' characteristics towards Thailand. There were two basic econometric models which were best fitted to the objectives for the analysis due to the primary survey data conducted earlier of the thesis academic procedures. Logit and Ordered Probit approaches were used to analyze and repeat experiments that involved in the huge number of times, variables of attitudes, statements of intentions, and willingness to build friendship between the dyad in the study.

3.3.1 Model 1 Logistic Regression

Peng, Lee, and Ingersoll, (2002) stated that logistic regression is best described and tested hypothesis about relationship between binary outcomes as one categorical variable and another predictor variable. When there are binary outcomes of interest, not continuous outcomes instead they are discrete values such as 1 or 0. Statistically, those outcomes are known as dichotomous variables, which Logistic Regression Model is the

most often used and more validity in econometric analysis. It links linear predictor to expected value of response. The probability Y outcome usually values as dummy variables with binary outputs, where Y takes on the value 0 and 1. The probability of Logit transformation as interval [0, 1] where the numbers could stand as the endogenous variable and the exogenous variables can stand between ∞ and -∞ for the estimation.

In this research, the conducted first objective is to analyze the factors on the attitude of Cambodians to make the decision to establish friendship with the Thais, for example, simplification of the statement in the questionnaire to Cambodians' respondents as "Would you like to make friends with the Thais? "Yes" to make friends, and "No" to make friends, indicated by 1 and 0. Logit model is used to best fit with objective primary objective in the research, testing on their decision to make friends with Thais.

Logit or binary logit regression is tested with dummy variable Y, which is supposed to observed and determined by Y^* as decision to make friend with Thais. There is two possible outcomes from the respondents, applied to estimate the probability of belongings groups of Yes and No. $Y_i = 1 \rightarrow$ Yes to make friends with the Thais and $Y_i = 0 \rightarrow$ No, otherwise. In a more mathematic form, we get as follow:

$$Y \begin{cases} 1 \text{ if } Y^* > 0 \\ 0, \text{ otherwise} \end{cases}$$

Where the index is close to positive, therefore we get Probability (Y=1) → 1, however, when the index is close to negative value; we get Probability (Y=1) → 0. Because the dichotomous outcomes are hard to describe with an ordinary least squares regression equation for two reasons: First, it is too extreme that linear trend will not appear. Secondly, error terms are not normal distributed through the entire data's range (Peng et al., 2002). Under perfect rationality, decision with greater expectation from respondents is positively chosen. In reality, there are some factors would limit decision makers from being perfectly rational. As a result, bounded rationality is being modeled to randomly estimate the binary outcomes comparison of decision makers. Hence, we can postulate the simple primary equation from the above postulation of Logistic Regression Model with dependent variable as below:

$$\ln\left(\frac{\text{prob Y}=1}{\text{prob Y}=0}\right) = \alpha + \sum_{k=1}^{n} \beta_k X_k + \varepsilon_k \qquad (3)$$

Where k=1, 2, 3,……., 415 respondents/ observations, β is the regression coefficient and X_k is a set of independent variables which considered as the elements to estimate the model from the survey such as gender, marital status, age, education level, occupation, income, travel experience, and their existing perceptions towards Thai people, and factors which are considered to influence on their decision to establish friendship with Tha is people. There are 13 factors in the category, and all in all there are 38 independent variables regressed, and the detail of the questionnaire is being listed in the appendix of the thesis. Table 1 give a list of variables with detail descriptions from the survey, which identifying dependent and independent variables. Those variables are statistically considered as random variables.

Table 3.1: List of independent variables for Logistic and Ordered Probit Models

Variables	Descriptions
Gender	1 for male, 0 for female
Married	1 for single and divorce, 0 is coded for married
Youth	Categorize from 0 year old to 24 years old
Adult	Categorize from 25 years old to 55 years old
Elder	Categorize from 56 years old and above
High school	Respondents with the highest level of education at high school degree
Bachelor	Respondents with the highest level of education at bachelor degree
Master	Respondents with the highest level of education at master degree
Doctorate	Respondents with the highest level of education at doctoral degree
Student	Current occupation of respondents as student
Freelance	Current occupation of respondents as Freelance
Private	Current occupation of respondents in the private companies
Government	Current occupation of respondents in the government sector
Lower middle income	Income rank of respondents is less than 500 US dollars per month
Upper middle income	Income rank of respondents is more than 500 but less than 1,000 US dollars per month
Rich	Income rank of respondents is more than 1,000 US dollars per month

Already have Thai friends	1=have Thai friends and 0 otherwise
Experience travel to Thai	1= have traveled to Thailand, and 0 otherwise
Week	Respondents have stayed in Thailand less than a week
Month	Respondents have stayed in Thailand from a week to a month
Year	Respondents who have stayed in Thailand from a month to a year
Many years	Respondents who have stayed in Thailand more than one years
Pleasure*	Travel purpose of respondents for pleasure
Education*	Travel purpose of respondents for education purpose
Business*	Travel purpose of respondents for business purpose
A**	I do not wish to make friends with Thais people
B**	I think Thais are not friendly toward Cambodians
C**	Thais seems to be terrible, cruel and unfriendly, and dishonest to Cambodians
D**	Thais are too conservative, they are too cold
E**	I find Thailand is my enemy because they want to make war with Cambodia
F**	The Thais are too open about gay and lesbian
G**	I have watched a lot of Thai's movies and I think Thais in real are like those in the movies
H**	Due to recent political issue between Cambodia and Thailand, I do not want to make friends with Thais
I**	There is a big gap in terms of living standard between Cambodians and Thais, thus the two countries can't live in peace together
J**	Due to Preah Vihear border conflict, Cambodia and Thailand should remain enemy as long as they could not find the solution
K**	The two countries should be ready to fight and compete economically in AEC
L**	Cambodia should not accept charity, donations, non-profit organization from Thailand
M**	There should be no peace between Cambodia and Thailand

* These variables are options questionnaires for those who never been to Thailand chose to skip them. For those who have experienced to visit Thailand, they could choose more than one option.
** These variables are 5-point likert scale statements of factors influence their decision and intention to establish relationship with Thais.

Outcome logistic model is given by the particular value of explanatory variables and unlike linear regression thus the model is used to derive with more meaningful explanation from the marginal effect model regression for the value changed in one particular indicator variable by holding other independent variables constant, on dummy dependent variable in logistic regression.

3.3.2 Model 2 Ordered Probit

Linear regression is not suitable when dependent variable is categorical according to a number of literature of previous studies (Alauddin and Tisdell, 2006). More importantly, when the survey is qualitative data where responses are coded as 1 is the worst and 5 is the best in the opinion survey. There will be bias from linear regression for the middle ranking such as 2, 3 and 4. Ordered categories or simply known as ordinal variables are commonly measured in various studies such as qualitative data including Likert Scale where respondents answer in the categories such as strongly disagree, disagree, neither agree nor disagree, agree, and strongly disagree and the ranking ordered with a statement. Scott (1997) supported that the interval between adjacent categories should be equal based on the implicit assumption in the sequential ordered regression models. It had been pioneered by McKelvey and Zavoina (1975) when the observed endogenous variable is ordinal. Some measured variables are ordinal and dichotomous, an appropriate and testable model for empirical study is to employ Ordered Probit estimation technique where variables are scaled as the function as a set of explanatory factors.

In order to find the determinants of the intentions to get involved with Thais, thus Ordered Probit model is used to fit with objective number 2, testing on the intention to get involved with the Thais by category. There are 9 intentions categorized in the questionnaires which have been distributed to respondents, which scaled as 5-point Likert

scale in the category and there are 4 thresholds over the latent variables where μ values represent the thresholds in the ranking. This view of variables is the best treated as the ordinal rather continuous variables (Winship and Mare, 1984). In a formal form of describing ordinal variables as endogenous variable, Y is denoted as observed ordinal variable $(-\infty < Y_i < \infty)$ and $\alpha_0, \alpha_1, \ldots \alpha_j, \alpha_{j-1}, \alpha_j$ denoted as cut-points or thresholds in Y distribution which provided incomplete information. From the latent regression, according to Briggs (2003), is considered primary equation of the Ordered Probit Model in the latent regression as the below as following:

$$y_i^* = \sum_{k=1}^{n} x_k \beta_k + \varepsilon_k \qquad (4)$$

Hence, $Y_i = j$ if $\alpha_{j-1} \leq Y_i^* < \alpha_j$ where J= 1, 2, 3,….,J), where J is ordered outcomes in the categories of Y, where there are 5 possible outcomes. Y_i^* is unobserved variable and roles as the predictor in the linear regression, the only observed variable is Y_i (Alauddin and Tisdell, 2006). Tactically, there are 9 categories of statements of intentions in the survey, Ordered Probit Model is being employed separately from intention 1 to intention 9, which are stated by N to V according to the questionnaires. Other 38 independent variables are used to estimate the same from the logistic regression model earlier. Those 9 statements have been asked to respondents to scale the statement such as: "I intend to visit Thailand once in my life." This method indicates the ranking response where participants can express their opinion based on the statements by scaling from 1 to 5 as the discrete outcomes: Strongly disagree coded as 1, disagree coded as 2, neutral coded as 3, agree coded as 4, and strongly agree coded as 5. Since Y^* is latent variable, it is not observable, the only observed variable is Y_i, and therefore we get in general:

$$y = 0 \text{ if } y^* \leq 0$$

$$y = 1 \text{ if } 0 < y^* \leq \mu_1$$

$$y = 2 \text{ if } \mu_1 < y^* \leq \mu_2$$

……

$$y = j \text{ if } y^* \geq \mu_{j-1}$$

In a more specific way of measurement in the model, we get as the following:

$$Y_i = \begin{cases} 1 \rightarrow \text{Strongly Disagree}, & \text{if } 0 < y_i^* < \mu_1, \\ 2 \rightarrow \text{Disagree} & \text{if } \mu_1 \leq y_i^* < \mu_2, \\ 3 \rightarrow \text{Neutral} & \text{if } \mu_2 \leq y_i^* < \mu_3, \\ 4 \rightarrow \text{Agree} & \text{if } \mu_3 \leq y_i^* < \mu_4, \\ 5 \rightarrow \text{Strongly Agree} & \text{if } \mu_4 \leq y_i^* < \mu_5 = \infty \end{cases}$$

$\mu_1, \mu_2, \ldots, \mu_5$ are unknown threshold values to be estimated with β which determine the degree of opinion for the respondents expected. In ordered Probit Model y_i^* = unobserved variables (represents Y, observed variable, as statements N until V based on the questionnaire, which are all in all 9 latent dependent variables stated for the intention statements of decision making to get involved with Thais). Ordered Probit from model 1 to model 9 are run separately of each statement with complementary of independent variables that could explain each dependent variable such as demographic factors, traveling experience and existing perceptions of Cambodians toward the Thais, and 13 factors which are the prominent variables that utilize the entire model to be regressed. In Ordered Probit Model, the inconsistency of parameters in the estimation causes the disturbance term ε_i from standard error of the non-normality.

The scale in each statement is very subjective, the spectrum of respondents' feelings proxy as the worst scale and 5 as the most positive and supportive opinion according to the statement. The outcomes are representing very close to the respondents with current perceptions of each question in the survey. Unobservable factors which respondents tend to unclearly express their opinion and the normal distribution with mean 0 and variance 1, are denoted by μ, This study, J=5, supposes there is the only interception of parameter

included. One comes out with probabilities, in general form, we get as the following:

$$\text{Prob}(y=j|x)=1-\Phi[\mu_{j-1}-\sum_{k=1}^{k}\beta_k x_k]$$

In a specific way in the probability of the equation, we get

$$\text{Prob}(y_i=1|x)=\Phi(\mu_1-x_k\beta_k)-\Phi(-x_k\beta_k)$$

$$\text{Prob}(y_i=2|x)=\Phi(\mu_2-x_k\beta_k)-\Phi(\mu_1-x_k\beta_k)$$

$$\text{Prob}(y_i=3|x)=\Phi(\mu_3-x_k\beta_k)-\Phi(\mu_2-x_k\beta_k)$$

$$\text{Prob}(y_i=4|x)=\Phi(\mu_4-x_k\beta_k)-\Phi(\mu_3-x_k\beta_k)$$

$$\text{Prob}(y_i=5|x)=\Phi(\mu_5-x_k\beta_k)-\Phi(\mu_4-x_k\beta_k)$$

All the probabilities have to be positive, therefore the value of μ are unknown threshold parameters of categories, where they must be in between

$$0<\mu_1<\mu_2<\ldots\ldots\ldots\ldots..<\mu_{j-1}$$

and Φ is the cumulative function of a normal distribution standard. Liao (1994) stated that there is the exception between the first and last categories, there is a difference between two adjacent accumulative probabilities which obtains the Prob(Y=J).

$$\text{Prob}(y\leq 1)=\text{Prob}(Y=1), \text{ and } \text{Prob}(Y\leq J)=1$$

Technically, β coefficients are not equal with the marginal effects of independent variables x in the estimation; plus they do not directly represent marginal effects of probability choice in the exogenous variables. For each endogenous variable of the intention statement of the marginal effects, in the survey will be separately regressed in order to well estimate with the outcomes due to independent variables as following:

$$\frac{\partial P(Y_i=1)}{\partial X_k}=\phi(\mu_1-\beta_k X_x)\beta_k$$

$$\frac{\partial P(Y_i=2)}{\partial X_k}=\phi(\mu_1-\beta_k X_x)-\phi(\mu_2-\beta_k X_x)\beta_k$$

$$\frac{\partial P(Y_i = 3)}{\partial X_k} = \phi(\mu_2 - \beta_k X_x) - \phi(\mu_3 - \beta_k X_x)\beta_k \qquad (5)$$

$$\frac{\partial P(Y_i = 4)}{\partial X_k} = \phi(\mu_4 - \beta_k X_x) - \phi(\mu_3 - \beta_k X_x)\beta_k$$

$$\frac{\partial P(Y_i = 5)}{\partial X_k} = \phi(\mu_4 - \beta_k X_k)\beta_k$$

The nature of categorical outcomes, the marginal effect of one variable is observed by the probabilities of alternative values of explanatory variables $X_{k.}$