

CHAPTER 4

Sample Design

4.1 Design of the Study

The research design is the framework that has been created to find answers to research questions here. A research design is the set of methods and procedures used in collecting and analyzing measures of the variables specified in this research problem or research study. The design of this research study defines the study type i.e. descriptive and fundamental study.

1. Identified the research problem clearly and justify its selection, particularly in relation to valid alternative design that has been used,
2. Reviewed and synthesized previously published literature associated with the research problem,
3. Clearly and explicitly specified research questions central to the problem,
4. Effectively described the data which will be necessary for an adequate testing of the hypotheses and explain how such data will be obtained, and
5. Described the methods of analysis to be applied to the data in determining whether or not the hypotheses are true or false.

4.2 Research Methodology

Globalisation and its impact on the courses are studied by collecting data on the most demanding STEM courses of various Indian as well as foreign universities based on subjective or purposive sampling. Online surveys are conducted in various departments of universities, considered to be popular among Indian students. Both teaching faculty and especially student community are to be covered during the survey. The analysis makes serious endeavour to unearth the causes behind enrolment pattern of Indian students in the U.S. At the same time the other countries attracting students from India in their tertiary education sector will find adequate attention in the present research.

In terms of the methods to be employed, the research will be based on extensive surveys, plus on interviews (with individuals and groups) along with desk research. The proposed key research instruments are to be standardised, closed questionnaires. The questionnaires will focus on diverse target groups: a questionnaire for students aiming to study abroad; a questionnaire for Indian students enrolled at higher education institutions of different countries (distributed personally and by e-mail); and a questionnaire for professors in higher education institutions. The survey targeted employees at various organisations

institutions selected for the survey. The initial sampling strategy involved compiling the names, addresses and e-mail IDs of potential participants who would be the sampling frame. The collection of the participant names and contact information depended to a great extent on the institutions and student unions. Given the limited time frame for the survey, the students abroad were surveyed by e-mail and also to forward the cover e-mail letter to their friends. So, this is a non-probability sampling method or —snowballll or —referrall sampling (Atkinson and Flint, 2001; Rea and Parker, 1997). This sampling is used as the size and distribution of the sample are not known with absolute certainty, and also the probability that the potential respondent selected as part of the sample is unknown. Besides, the participatory surveys were conducted in India. Since the scope and extent of such surveys are limited, non-participatory surveys were also conducted simultaneously.

The ordered response model has assumed that the explanatory variables have the same impact across each and every category of the dependent variable – is called the —parallel regression assumptionll (Long and Freese, 2001). In the generalized ordered models, a separate parameter vector was estimated for each of the J categories (e.g., $\beta^1, \beta^2, \dots, \beta^J$). The parallel regression assumption is tested with an approximate LR test and a Wald test (Long and Freese, 2001) and it was rejected by the researcher.

Even if the parallel regression assumption was violated, the research is based on the ordered probit model since their shortcomings outweighed their advantages. The multinomial logit model rarely preserves the inherent ordering of return intention categories and does not incorporate this information during estimation of the coefficients of the explanatory variables. This causes a loss in the efficiency of the estimators (Long, 1997). As the —definitely not returnll category has relatively few observations, it shows more intense feeling about returning than —unlikely to returnll category. So, the results are presented from the ordered probit model. A bigger sample size and fewer explanatory variables would make the generalized models more feasible. The model selection is based on the estimation of exploratory ordered probit models. The ordered probit model was taken as the estimation method based on the characteristics of the dependent variable. The procedures of model selection are used to determine the set of regressors to ascertain in the final estimation model. The set of possible regressors do not have the same number of valid points (cross-sections) due to missing responses. After determining the initial set of explanatory variables the next stage in model selection might be complicated by non-linearities plus interactions among the regressors. A saturated model incorporates all probable variables, interactions as well as

higher-order terms using a backward elimination procedure. At each step, terms that are statistically insignificant individually and also do not contribute significantly to the fit of the model are excluded. The elimination procedures are continued until further model reduction involves a significant deterioration in model fit. The results of the study are derived from data from survey collected by the author. In the econometric analysis of return intentions, the emphasis is on determining the factors important in explaining the return intentions of engineers. The dependent variable is the likelihood of returning to India rooted in the response to the question —What about your current intentions about returning to India? The following table reveals the possibilities presented to respondents.

Dependent Variable, Return Intentions

Response Categories	Index
Students	
I will return as soon as possible whenever I get a job in India.	1
I will return immediately after getting a job in India.	2
I will definitely return but not soon even if I get a job in India.	3
I will probably return to India.	4
I don't think that I will be returning.	5
I will definitely not return.	6

These choices comprise a set of ordered categories where each consecutive category points to an increase in intensity in the employees' intentions to stay in the U.S. The categories with a higher index value show stronger feelings about not returning (staying) to India. This suggests that positive coefficients on the independent variables indicate an increase in the probability of having non-return intentions, whereas the negative coefficients point to an increase in the probability of having return intentions. The change in intensity between categories can hardly be assumed to be uniform. The observed discrete index is given by

$$y_i = \{1, 2, 3, \dots, J\} \quad (1)$$

where i is the observations and J is the number of categories of the dependent variable. The assumption is that a continuous and latent variable underlies the discrete, ordered categories may be explained by a set of observed characteristics and a random element as given below:

$$Y_i^* = \beta'X_i + u_i \quad (2)$$

As y^* is the unobserved —return intention variable, X is taken as the $(k \times 1)$ vector of explanatory variables, β is considered as the parameter vector to be estimated and u is taken as the random disturbance term. The relationship is shown between discrete, observed y and unobserved, continuous y^* is mentioned below:

$$y_i = \begin{cases} 1 & \text{if } Y_i^* \geq \mu_1 \\ \mu_2 & \text{if } \mu_1 < Y_i^* < \mu_2 \\ \mu_3 & \text{if } \mu_2 < Y_i^* < \mu_3 \\ \dots & \dots \\ \mu_J & \text{if } \mu_{J-1} < Y_i^* < \mu_J \\ J & \text{if } Y_i^* \geq \mu_J \end{cases}$$

as $\mu_1, \mu_2, \mu_3 \dots \mu_{J-1}$ are threshold parameters – linking y to y^* estimated along with explanatory variable coefficients. The ordered probit specification assumes an (underlying) normal distribution for the error term and is used in this research to estimate the model of return intentions. According to an ordered probit specification, the probability of an observed response falling into an arbitrary category j is mentioned below as:

$$\text{Prob } i = j = \frac{\Phi(\mu_j - \beta'x_i) - \Phi(\mu_{j-1} - \beta'x_i)}{\Phi(\mu_j - \beta'x_i)}$$

where Φ is the cumulative normal distribution.

Distinguishing this probability in relation to the explanatory variables renders the marginal effect of each on the probability of choosing category j . Model estimation has been carried out by maximum likelihood (ML) estimation technique as it has been shown that ML produces the unbiased and efficient estimates, especially for nonlinear models. As part of the non-participatory survey, an e-mail cover letter was mailed to potential participants identified through the internet search process. Employees were invited to take part in the survey and to forward the mail to others who they felt would fit the potential survey population.

The survey results are not to be generalized. Rather the combination of participatory observation and —snowball sampling resulted in responses from science, mathematics and technology students and faculties in India as well as the United States.

The empirical specification of the model can be found in an ordered probit model – estimated about return intentions of engineers where the return intentions are the dependent variable discussed in the previous section in the six categories from values 1 through 6. The wage differences existing between the home country and the country of destination is viewed as the

primary factor of migration flows. The push-pull factors are constructed as dummy variables assuming the value —one if they score high on a 5-point Likert scale and value of —zero if they were relatively less important in return intentions, i.e. received a score of —three or less. Push factors emanate from the ambiance of the home country that prompt people to migrate. Various institutional, professional, economic and political factors, and also unsatisfactory income, inadequate work environment, absence of career development opportunities in addition to economic and political uncertainty are guiding factors in brain drain. Pull factors, are the opportunities offered by the host country e.g. better income, better learning as well as professional development prospects for attracting the skilled human resources to the host city. To explain the return intentions several —subjective variables are applied to determine the significance of a range of social, economic, political and human capital factors. The empirical model is chosen based on goodness-of-fit statistics; mainly the AIC and McFadden’s adjusted R. The likelihood ratio test was used too when comparing nested models. So, these three statistics produced similar results. The details of the model are supplied in Appendix A.3. The next section intends to discuss the rationale for considering each variable and their empirical significance of engineers in determining the return intentions/willing to settle in Kolkata. The push-pull factors are dummy variables taking on the value of —one if they have scored high on 5-point Likert scale (e.g., received either —five or —four) and a value of —zero if they were comparatively less important in determining return intentions (e.g., received score of —three or less) (Güngör, et al., 2007). The paper attempts to define the change of intention for showing the difference between a student’s initial and current return intention. The target population is the Indian students studying in the Masters in STEM and medicine courses in selected top 20 institutions of United States. The questionnaire focuses on perception related variables covering various issues ranging from family, career, financial issues and so on. The web based surveys were conducted with the help of Indian student associations and snowball survey methods during February 2010 to July 2016. About 3.3 percent population was responded in the survey. In table 4.1, a set of education related variables will attempt to focus on the effects of how the level as well as subject of study may cause change in return intention.

Table 4.1 Variable Description and Coding Scheme

Variable description and coding scheme	
Set 1 : Demographic and family related variables	
Variable	
Age	Age
Years of stay in the U.S.	Duration of stay in the U.S.
Work experience	Years of work experience prior to study
Single or married	1 for single, 0 otherwise
Family backs migration plan	1 if family supports non-return, 0 otherwise
Father tertiary educated	1 if father tertiary educated, 0 otherwise
Set 2: Education related variables	
PhD	1 if PhD, 0 otherwise
STEM	1 if STEM, 0 otherwise
Medicine	1 if medicine, 0 otherwise
OPT opportunities after education	1 if OPT, 0 otherwise
Have had education other than USA and India	1 if education received, 0 otherwise
Scholarship opportunities	1 if Scholarship, 0 otherwise
Set 3: Home perception related variables	
Better work opportunities	1 if work opportunities, 0 otherwise
Competitive wage at India	1 if wage competitive, 0 otherwise
Good use of skill at India	1 if skill use, 0 otherwise
Good lifestyle in India	1 if good lifestyle, 0 otherwise
Close ties at home country	1 if close ties, 0 otherwise
Better work opportunities work opportunity with specific faculty	1 if work opportunities, 0 otherwise

A proxy of geographical mobility has also been included. The variables are discussed below. By the latent variable approach, the bivariate probit model is mentioned as below:

$$Y^*_1 = X\beta_1 + \varepsilon_1$$

$$Y^*_2 = X\beta_2 + \varepsilon_2$$

$$E[\varepsilon_1 | X] = E[\varepsilon_2 | X] = 0$$

$$Var[\varepsilon_1 | X] = Var[\varepsilon_2 | X] = 1$$

$$Cov[\varepsilon_1, \varepsilon_2 | X] = \rho,$$

As Y_1^* has latent propensity to change one's intention, Y_2^* has latent propensity to return, and X is the vector of explanatory variable. Y_1 is 1, in case there is change of intention ($Y_1^* > 0$) and 0 otherwise, and Y_2 is 1 if the present intention is to come back ($Y_2^* > 0$) and 0 otherwise. In table 2 the breakdown of explanatory variable has been shown by the dichotomous outcome variable; that means whether the intention of returning has changed or not (Soon, 2010).

Data is to be collected from various sources:

Secondary data from

- a. Secondary data published by the Institute of International Education (IIE) with support from the State Department's Bureau of Educational and Cultural Affairs (United States) and the U.S. State Department.
- b. CGS International Graduate Admissions Surveys

Primary data from

- a. Field survey data based on survey among students in selected departments of Indian Universities/Institutions.
- b. Online survey of scholars selected departments of sample U.S. universities.

4.3 Population of the study

The scholars belonging to different science and technology streams participated in the survey. Out of 1752 students in the Indian students' category studying in the sample institutes in India, 465 responded. Another survey was conducted among the faculty in those institutes. The respondents were 89. The faculty members also belong to the same departments as the students. Only 18 percent of the faculty members responded in a closed ended questionnaire. The Indian students studying in the U.S. were also surveyed by online questionnaires. The surveyed institutions are California Institute of Technology, Missouri University of Science and Technology, Massachusetts Institute of Technology, Michigan Technological University, Carnegie Mellon University, Cornell University, North Carolina State University – Raleigh, Princeton University. The STEM students in postgraduate and above were surveyed. The sample population is 2647.

4.4 Sample and Sampling Technique

Stratified random sampling technique has been used for calculation of the sample. The technique involves dividing the sample population into smaller groups called strata. In

this case the strata are formed based on members' shared attributes or characteristics. This technique is useful in this study because of variety of attributes of sample selected here. Hence, at least one sample is selected for fair representation even if probability of it being selected is 0. The STEM subjects are equally represented at the time of selection of sample. The U.S. government approved list of STEM subjects are considered for sample selection. The broad groups are taken into consideration. The samples are taken from broad groups of subjects.

Box: 4.1 STEM Subjects taught in United States

The major subjects in STEM list as listed in ICE.gov are agro-ecology and sustainable agriculture, computer and information science, engineering, biological science, mathematics, air science, air power studies, naval science, intelligence/command control/information operations others, behavioural science, physics, chemistry, earth science, psychology, computerised forensics, archaeology, econometrics, clinical laboratory science, management science and quantitative methods etc.

4.5 Tools of the Study

The available tools for testing the hypotheses are:

1. Questionnaires
2. Schedules and
3. Observation techniques.

Questionnaires were distributed and data were collected personally in the institutions located in India; for abroad, the online questionnaires helped in collection of data.

The schedule was presented by the researcher. The questions were asked and the answers were noted down by the researcher. The list of questions was a mere formal document. It has aided in delimiting the scope of the study and to focus on the circumscribed elements needed for the analysis. It also helped delimiting the subject. For example questions were asked about their intention to migrate to the U.S. for higher education and the possible causes. The expatriate students in the U.S. were also asked about their return intentions and

the factors governing their decisions to stay or go back to India. For example, income, career prospects, job opportunities, meeting with family, cultural factors or contribution to mother land were major decisive factors.

Observation technique was used for evaluation research. It is useful for evaluating cognitive and non-cognitive aspects of the respondent. It is applied in evaluation performance, attitudes, interests, values towards their problems. The observation method was applied to the respondents in Indian institutions for assessing subjective issues in migration decisions. There are emotional issues attached to families, friends and place of origin of a person which can be assessed by talking to the respondents and getting responses rather than by filling structured questions only.

4.6 Procedure of Data Collection

Primary data were collected by a combination of participatory and non-participatory surveys. Both closed ended and open-ended questions were included in the questions. The domestic surveys were conducted by the participatory method. The questionnaires were distributed among the respondents. Both students and faculty were included. The international surveys were conducted by sending emails to the Indian students' associations of the respective institutions. Since all the contact lists of respondents were not accessible, snowball sampling techniques were used in order to get access to rest of the participants. Many emails were not answered, even if the available responses were sufficient for data analysis. Secondary data source includes data published by the Institute of International Education and Times Higher Education Ranking.