# Three Essays on Labor Markets of South Asian Countries

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Economics

By

Mrittika Shamsuddin, M.A.

Washington, DC
May 12, 2013

**Three Essays on Labor Markets of South Asian Countries**

Mrittika Shamsuddin, M.A.

Dissertation Advisors: James Albrecht and Susan Vroman

Abstract

South Asia has been experiencing robust economic growth of about 6% over the last decade. However, it is also the home to about 44% of the world's poor. About 571 million people of the region live under $1.25 a day. About 80% of its labor force works in the informal sector with low skill and low earnings. It has the largest working age population (56%) and the lowest female labor force participation (18%). It is also the largest sender of emigrants to the world. In order to promote sustainable and inclusive growth, South Asia needs to find productive employment for its key assets, which are its people. The governments of South Asian countries have taken many initiatives in this regard including education programs like the female secondary school stipend program, employment programs like the food for work and the employment generation for the poor and demand side interventions like enhancing firm's access to electricity through subsidy. This dissertation looks into the labor market effects of two such programs and whether international migration is an economically advantageous outside option for the emigrants from South Asia.

The first chapter studies the long-term effects of Bangladesh's female secondary school stipend program on both educational outcomes and economic empowerment of women and finds that the program can be associated with an increase in education level completed by at least 0.21 years, labor force participation by 2 percentage points and with an increase in the likelihood of women working in the formal sector by 3 percentage points. Female wages also increased, reflecting increased productivity. The second chapter finds that migrants from Bangladesh go to the country where they earn the most and the relative stock of more-educated Bangladeshi migrants is higher in the destination where the earnings difference between the skilled and the unskilled workers is higher. The third chapter builds a three-sector model of the Indian labor market, calibrates it using pre-MNREGA data and finds that providing manual employment for a fixed number of days like the MNREGA in India can improve welfare for the unskilled but may have adverse effect on the formal sector and skilled workers.

INDEX WORDS: female secondary stipend program, employment program, migration, sorting, selection, informal, cash-transfer, labor markets

# Dedication

This thesis is dedicated to my family for their continuous support.

## Acknowledgments

I would like to thank Professor Albrecht, Professor Vroman, Professor Jack and many others who helped me along the way with their insightful comments and suggestions.

Table of Contents

# Chapter 1

# Labor Market Effects of Female Stipend Program in Bangladesh

## 1.1 Introduction

The World Economic Forum Report on Global Gender Gap (2012) ranked Bangladesh 86th according to its gender gap index. The gender gap in Bangladesh is lower than in both neighboring India and Pakistan although GDP per capita in these countries is higher than in Bangladesh. The report based its ranking on education, economic opportunity, health and political empowerment as Table 1.1 shows. It ranks Bangladesh on top of most South Asian and all Muslim countries. It also reports that the gender disparity in Bangladesh has decreased drastically since the mid nineties. Figure 1.1 shows the trends in labor force participation and the daily real wage rate using various issues of the Household Income and Expenditure Surveys of Bangladesh (HIES)[1]. Since 1995, female labor force participation has increased from 10% to 18% while male labor force participation has been around 90%. Between 1995 and 2010, the female wage rate increased by 7.6 times while the male wage rate increased by 2 times. Although, labor force participation of women and their wage rates are still substantially lower than those of men in Bangladesh, girls surpass boys in school enrollment, especially in secondary school (World Bank, 2012). This paper aims to study whether there is a policy intervention that has contributed to this extraordinary performance.

[1] Author's calculation from various surveys, including wages from salaried and day laborers.

Table 1.1: World Economic Forum on Gender Gap Ranking

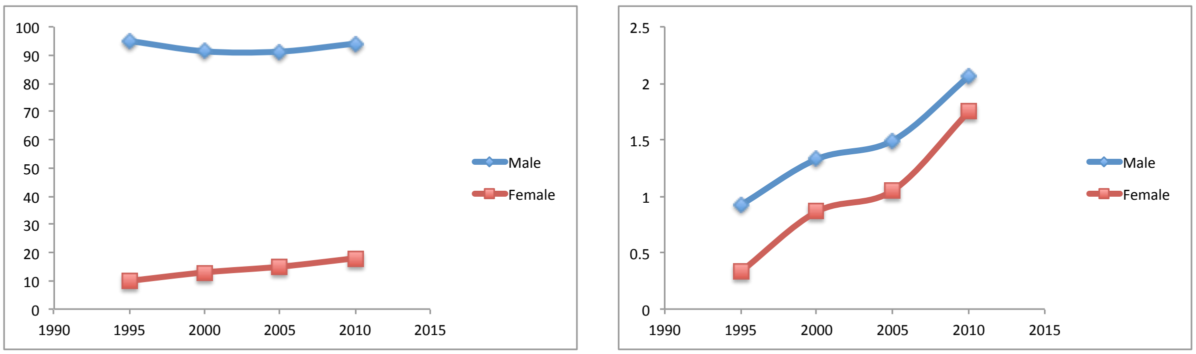
| | Bangladesh | India | Pakistan |
|---|---|---|---|
| Overall | 86 | 105 | 134 |
| Economic Opportunity | 121 | 123 | 134 |
| Education | 118 | 121 | 129 |
| Health | 123 | 134 | 123 |
| Political Empowerment | 8 | 17 | 52 |

Bangladesh may be the first country to give girls who are attending secondary school stipends, tuition fees, books and exam allowances. Secondary school refers to grade 6 to grade 10 in Bangladesh. The Female Secondary School Assistance Project (FSSAP) was jointly initiated by the World Bank, the Asian Development Bank and the Government of Bangladesh in 1994. It covers 57 out of 64 districts and 95 percent of all thanas (sub-districts). The program aims to empower women by fostering education so that women can participate in economic and social development of the country.

Khandker et al. (2003) and Fuwa (2001) find that 8 percent of the increase in school enrollment can be contributed to FSSAP. Following Bangladesh, countries like Pakistan, Brazil, Mexico, Nicaragua and Paraguay have taken similar interventions, though on a smaller scale, and research shows that in all these countries, secondary school assistantship programs increase school enrollment and attendance. To the best of my knowledge, there has not been any study investigating the labor market effects of this policy. This paper extends the literature by studying the long-term effects of FSSAP on the labor market of Bangladesh. The effect is identified through variation in exposure by gender, location and age. Using regression discontinuity methods in a triple difference and quadruple difference framework, as used by Oreopoulos (2006), this paper studies the effect of FSSAP on female labor force participation, wages, the likelihood of working in agriculture, the likelihood of working for the government and the likelihood of working in formal sector. This paper also extends the work of Duflo (2001) who study the labor market consequences of a school construction program in Indonesia, by studying the effects of a stipend program targeted to girls in Bangladesh.

The rest of the paper proceeds as follows. Section 2 describes FSSAP in detail. Section 3 gives the literature review while Section 4 describes the methodology and the data. Section 5 reports the results and Section 6 concludes.

Figure 1.1: Labor Market




(a) Labor Force Participation (b) Daily Real Wage

## 1.2 The Female Secondary School Assistantship Project

FSSAP was initiated in 1994 to provide female students with tuition-free education, book allowances and a small stipend. Each girl received tuition-free education, a yearly book allowance of Tk. 250 ($4), a monthly stipend of Tk. 435 ($7) and a one-time SSC exam fee of Tk. 250 ($4). It was jointly financed by the Government of Bangladesh, the World Bank, the Asian Development Bank and the Norwegian Agency for Development Cooperation. Grades 6 to 10 constitute secondary school in Bangladesh. Any girl enrolled in a secondary school which was located in the areas covered by the program was eligible as long as she remained unmarried, attended 75 percent of the school days and scored 45 percent in the school-level exams. The stipend was paid directly to an account in the girl's name in the nearest Agrani Bank, a state commercial bank with branches all over Bangladesh. The stipend program was complemented with curriculum reforms, occupational skill development programs, instructional materials development, teacher training, recruitment of female teachers, improvement of school infrastructure, awareness programs at the community level, and institutional capacity building. FSSAP's aim was to increase female access to secondary education, reduce the fertility rate by delaying marriage and improve women′s economic and social empowerment.

FSSAP was administered in 57 out of the 64 districts and covered 24950 schools and madrasahs (Islamic schools) out of the total of 28140 schools in 2002 (Banbeis, 2012). The number of beneficiaries ranged from 1.5 million to 2.8 millions. The list of districts covered is given in the data appendix (Table A.1). Details about the stipend program are also given the data appendix (Table A.2). The project document claims that it is a nationwide program phased in randomly however 7 districts were left out without any explicit reason. 3 of these districts were metropolitan areas while the other four were rural areas (Schurmann, 2009).

Figure 1.2: Gender Differences in Secondary and Tertiary enrollment


Source: World Development Indicators

Secondary school enrollment for both boys and girls has been increasing since 1990. Figure 1.2 shows that between 1991 and 1998, there has been a dramatic jump in secondary school enrollment for both boys and girls. Although secondary enrollment increased by 40 percentage points for girls, it increased only by 20 percentage points for boys. More importantly, since 1998, female enrollment has surpassed male enrollment in secondary school. According to the figure, tertiary enrollment increased but almost identically for both males and females. This raises interest in the role of FSSAP in the drastic increase in female enrollment rate since it benefitted only the girls in secondary school. However, data is not available from 1993 to 1997 in the World Development Indicators. Consequently, we cannot attribute the increase to any intervention since we do not know when exactly the jump occurred. Using Household Income and Expenditure Survey data (HIES), I find that between 1991 and 1995 secondary school enrollment among girls increase by about 50% while among boys, it increased by 30%. Figure A.1 shows that the fall in enrollment from primary to secondary school among girls decreased substantially between 1991 and 1995 while that among boys remained the same. Khandker et al. (2003) use school-level data from 1991 to 1998 to study the impact of FSSAP on female enrollment. Using variation in the introduction of the program to different class cohorts, they find that an additional year of program exposure increases female enrollment rate by 8 percent.

### 1.3 Literature Review

Although there have been plenty of research on the role of education and education policies on labor market outcomes, there has been no research done on the effect of FSSAP on average years of completed education and on the labor market in Bangladesh. Duflo (2001) evaluates the effects of building schools on education and earnings in Indonesia using a single cross-sectional data. Combining differences across regions in the number of schools constructed with the differences across cohorts induced by the timing of the program, she finds that each primary school constructed per 1000 children lead to an increase of 0.12 to 0.19 years of education and 1.5 to 2.5 percent increase in wages. I use repeated cross-sectional data to study the labor market effects of a secondary school stipend program using differences in exposure by regions, gender and birth-cohorts.

Heath and Mobarak (2012) study the effects of the explosive growth in the garment sector on young girls' enrollment and find no significant effect of the female schooling subsidy on enrollment. They conclude that demand from the garment sector plays a key role in the enrollment decision. They use recall data from only four sub-districts in Dhaka division, where most of the garment sector firms are located. It is unlikely that the same result will hold for all of Bangladesh. Recall data may be the reason they find similar discontinuities in the years just before and after 1994 since people may not remember exactly in which year their children went to which grade.

FSSAP is a conditional cash transfer and research done on other countries finds significant impacts of cash transfers on school enrollment and attainment, child labor, current consumption, and on the marriage and fertility decisions of adolescents. Alam et al. (2011), Behrman et al. (2005), Maluccio and Flores (2005) and Soares et al. (2008) report these results for Pakistan, Mexico, Nicaragua and Paraguay respectively. However, none of these papers look at the longer-term impacts of these programs. Since it has been almost 18 years since FSSAP has been implemented, the first beneficiaries are now in the job market, and this provides a unique opportunity to study the long-term labor market outcomes of such transfers. This paper attempts to study the role played by FSSAP on the increase in female labor force participation and their wages. It uses regression discontinuity with triple differences along with quadruple differences to study the impacts of FSSAP. In this regard, my estimation method is similar to Oreopoulos (2006) and Harmon and Walker (1995). More education seems to increase productivity and hence wages and increases the likelihood of women working in income generating activities.

## 1.4 Methodology and Data

### 1.4.1 Methodology

The date of birth, gender and the region of birth jointly determine an individual′s exposure of the program. Bangladeshi children normally attend secondary school between the ages of 12 to 16. All children born in 1977 or before 1977 were 17 or older when the program was implemented in 1994 and did not benefit from the program.

It affected only the relatively young female cohort among those currently in the labor force, those between age 25 and 32 years in 2010, living in the 57 districts covered under the program. In an ideal world of true counterfactual impact analysis, the net impact of the stipend program would be the difference between the relevant outcome indicator in presence of the intervention and that in absence of the intervention. However, one never observes the unit of evaluation in both states. A naive method to address such missing data problem is to focus on changes in outcome variables over time of the treated unit. This estimator can overestimate the impact as the outcome variable might have changed even without the program.

One approach is to measure the changes in the outcome variables in both the treated and control group over time and take the difference in changes between the two groups. This difference-in-difference estimator takes care of the time-trend sensitivity present in before-after estimator by sweeping out time-invariant effects on the outcome variable. There are four potential control groups I can use: females of a certain age in non-treated areas, older females in the treated areas, males of that age group in treated areas and males of that age group in non-treated areas. I use three of the groups in difference-in-difference-in-difference estimation (D3) and all the four groups in a difference-in-difference-in-difference-in-difference estimation (D4).

The crucial identification assumption in difference-in-difference is that outcomes in treated and control groups would have followed a parallel path in the absence of the program. However, this assumption is unlikely to be satisfied if we take females in the non-treated areas as the control group or the males in the treated areas. The fact that the program was implemented in only certain districts and was targeted to females may introduce some endogeneity problems. To deal with this, I choose males and older females in the treated areas as my control group for the D3 estimation. This should alleviate the problem of initial district conditions that affect subsequent growth. Having the treatment and the control groups from the same area makes them comparable from this point of view. The specification of the D3 model is

$$Y_{it} = a_0 + a_1 S_i + a_2 T + a_3 f + a_4 S_i * f + a_5 f * T + a_6 S_i * T + a_7 S_i * T * f + a_8 C_{it} + a_9 e \quad (1.1)$$

where $Y_{it}$ is the outcome of interest, f is the dummy for gender, $C_{it}$ is the vector of covariates, $S_i$ takes the value of 1 for the age group that is supposed to get the policy and 0 otherwise and T is the time period dummy which is equal to 1 if the year is after the implementation of the policy and 0 before the implementation. The coefficient $a_7$ measures the net impact of the program.

The D4 estimator compares relative female outcome variables for different age groups in treated areas relative to non-treated areas with the relative male outcome variables for different age groups in treated relative to non-treated areas. It separates out some of the biases from the differential growth effects that arise due to different initial conditions. It also takes into account alternate development programs as long as they are not gender targeted. In Bangladesh, no such program is in place that targets gender except the FSSAP. The regression framework is then:

$$Y_{it} = u_0 + u_1 S_i + u_2 f + u_3 S_i * f + u_4 T + u_5 p + u_6 p * T + u_7 S_i * p + u_8 S_i * T + u_9 f * T$$
$$u_{10} p * f + u_{11} S_i * T * f + u_{12} S_i * p * f + u_{13} S_i * p * T + u_{14} S_i * p * T * f + u_{15} C_{it} + u_{16} e \quad (1.2)$$

where p is a dummy for treated areas and $u_{14}$ is the D4 impact estimate.

The third way to deal with selection bias is to use a regression discontinuity (RD) technique that exploits the fact that the program was given to only those who were at secondary school. Girls who were 16 years old at 1994 were the first graduates from the program. So, girls who were born before 1978 in the treated areas were not affected by the program while girls born on or after 1978 were eligible for the program for at least a certain portion of their secondary education. 1978 is the cut-off birth-cohorts. The regression discontinuity design identifies and estimates the program impact in the neighborhood of the cut-off point for selection into the program. The birth-cohorts within the small interval of the cut-off should be similar in the presence or absence of the program and thus provide good comparison groups. In the regression discontinuity literature, there are two broad methods to calculate the net impact: nonparametric and parametric. While nonparametric approach avoids any misspecification in that of the parametric methods, it effectively uses a smaller sample of observations close to the cut-off point. Parametric specification on the other hand extrapolates from above and below the cut-off point to all observations in the data. Following Imbens and Lemieux (2007), who argue that local linear regression is the preferred way to do regression discontinuity, I use local linear regression for most of my analysis. I use a rectangular kernel and optimal bandwidths as Imbens and Lemieux (2007) suggests. Using a triangular kernel does not change the results significantly. Following Oreopoulos (2006), I also regress the dependent variables on a control for a birth-cohort quartic polynomial and an indicator for whether the cohort received free secondary education for some fraction of their secondary school in the treated areas.

$$Y_c = b_0 + b_1 1(YOB \geq 1978)_c + g(YOB) + v_c \tag{1.3}$$

where $Y_c$ is the outcome of interest for cohort c, YOB is the year of birth. $b_1$ is the consistent estimate of the program impact. The quadratic polynomial, $g$, allows for heterogeneous impacts across different birth-cohorts. In line with Oreopoulos (2006), I also estimate the D3 and the D4 estimator within the regression discontinuity design by adding a control function for the birth-cohorts in the D3 and in the D4 framework. The D3 estimator equivalence in regression discontinuity framework becomes:

$$Y_{cdt} = a_0 + a_1 X_c + a_2 f + a_3 X_c * f + a_4 C_{dc} + g(YOB) + e_t + e_d + e_{ctd} \tag{1.4}$$

where $Y_{cdt}$ is the outcome variable of cohort c in district d from survey year t, $C_{dc}$ is a vector of district level controls while $e_t$ is the fixed effects for different survey years, $e_d$ is the fixed effects for each district and $X_c$ is the indicator variable, $1(YOB \geq 1978)_c$. $a_3$ is the program impact. The D4 estimator equivalent in the regression discontinuity framework becomes:

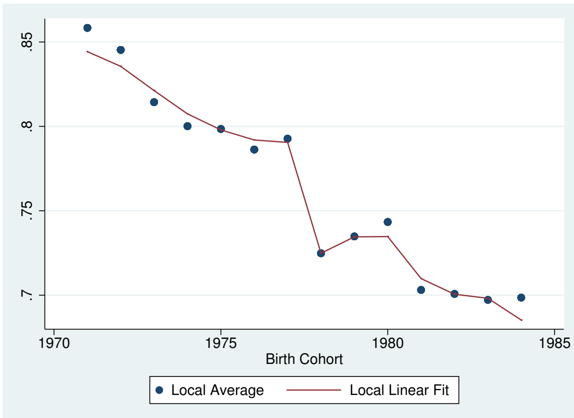
$$Y_{cdt} = u_0 + u_1 X_c + u_2 f + u_3 X_c * f + u_4 p + u_5 X_c * p + u_6 p * f \\ + u_7 X_c * p * f + u_8 C_{dc} + g(YOB) + e_t + e_d + e_{ctd} \tag{1.5}$$

where $u_7$ is the net program impact and p is the dummy for treated areas.

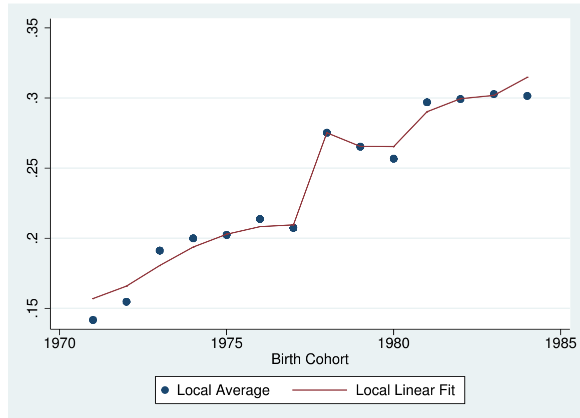
### 1.4.2 Data

The data used for the analysis is derived from combining 4 quinquennial Household Income and Expenditure Surveys (HIES) from 1995 to 2010 (1995, 2000, 2005 and 2010). The combined dataset contains 71942 females who were of age 25 to 50 years between 1995 and 2010 but only the ones aged between 25 and 32 years in 2010 and between 25 and 27 years in 2005 attended secondary school after the policy was implemented. The sample size for each birth-cohorts is documented in Table A.3 of the appendix. Assuming that girls attend secondary school from age 12 to age 16, girls born from 1978 to 1981 got the stipend for some fraction of their time in school while girls born in and after 1982 got all five years of their secondary school funded by the policy. According to the project documents from the Education Ministry of Bangladesh, the program was implemented in all the districts by the end of 1994. All data are collapsed into cell means by birth years, survey years, districts, and gender and weighted by population. I code the schooling variable for individuals as the highest grade completed capped at 12 years to impose a uniform top code across surveys. Figure 1.3 shows that the fractions of the birth-cohorts who left school after completing primary school has decreased substantially for the 1978 cohort while the fractions of the birth-cohorts who left school after completing secondary school increased significantly by 7 percentage points.

Figure 1.3: Female Primary and Secondary completion rate




(a) Fraction with completed primary only




(b) Fraction with completed secondary only

I create an indicator variable equal to one if the person is born in or after 1978. It includes all females who got the subsidy for one or more years. I create another indicator variable, young, which includes people born after 1982, including females who got the subsidy for all five years of their secondary education. The third indicator variable, middle, includes females who got some subsidy but for less than five years. Another indicator variable for the treated regions is created. The monthly earning variable is calculated by multiplying days worked in a month with the daily wage or the monthly salary reported depending on what is recorded in the survey. All wage data are adjusted for inflation by the consumer price index (CPI) obtained from the World Bank Development Index. The labor force participation rate includes those who were employed or looked for a job in the week before the interview for the survey or who worked for at least an hour during the previous month. Informal sector employment consists of people working as self-employed or as day laborers while the formal sector employment consists of regular employees. I match each individual with the characteristics of the districts in which he or she lived. The regional control variables that I use include the fraction of the population in the district living in rural areas, the illiteracy rate in the district and the fraction of the people in the district working in the manufacturing sector.

Figure 1.4: Average years of completed education by birth-cohorts

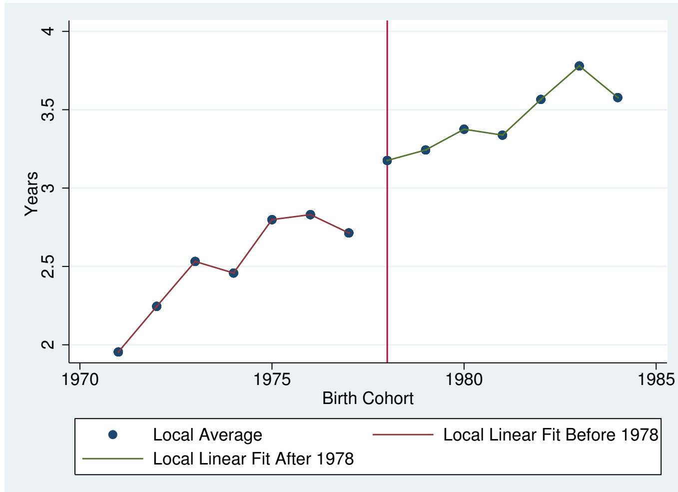



Figure 1.4 shows that the average education attainment of females born after 1978 increased by about half a grade in treated areas. The policy is associated with an increase in labor force participation of about 8 percentage points and with an increase in earnings by 55 percent in a year (Figure 1.5 and Figure 1.6) of the same female cohort. The increase in earnings provides evidence of the increase in productivity of the girls due to increase in their educational attainments. The sample size is restricted to only females in the treated regions. The fitted lines in Figure 1.4, 1.5 and 1.6 are obtained by local linear regression. I do not see any similar discontinuity in male education attainment, male earnings or male labor force participation as Figure 1.7 and Figure 1.8 reveal.

Figure 1.5: Average female labor force participation by birth-cohorts

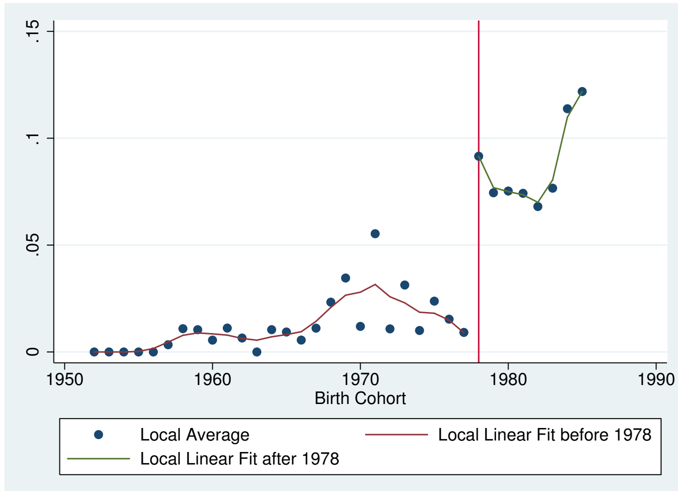
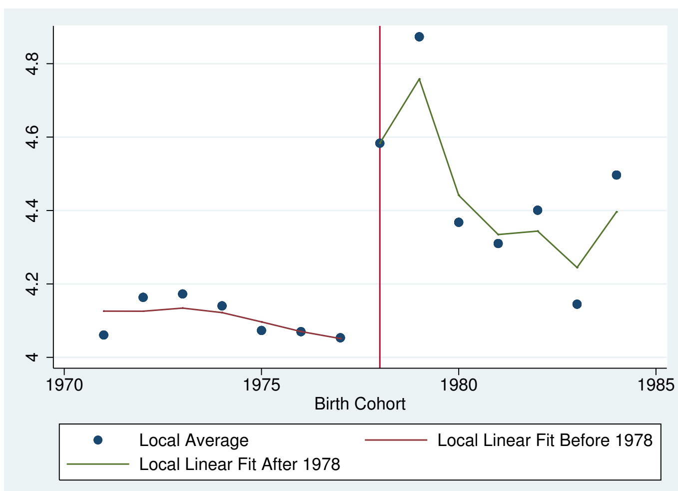



Figure 1.6: Average female log real monthly earnings by birth-cohorts




For the D3 and the D4 analysis, I only use individual level data from the 1995 and 2010 household surveys. In 2010, all females in the age range of 25 years to 32 years of the 57 districts received FSSAP while in 1995 no one in the age range of 25 to 50 years received FSSAP. In the 2010 survey, the cohort aged between 25 and 27 years got FSSAP for all 5 years of their secondary school while the cohort aged between 28 and 32 years got the treatment only for a fraction of their secondary school years.

Figure 1.7: Average years of completed education by birth-cohorts for male

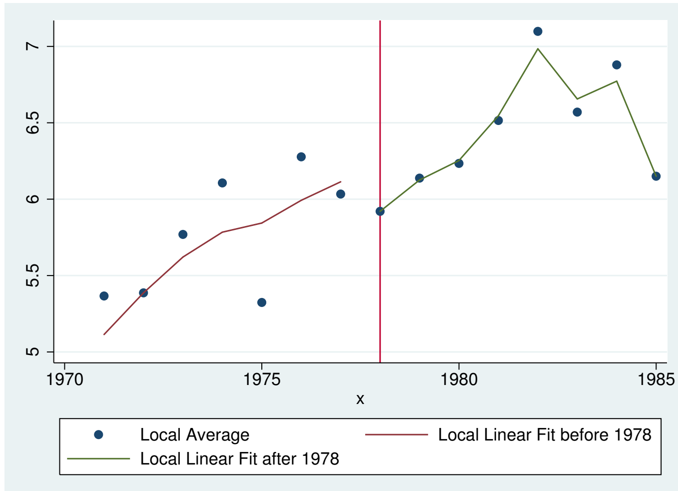
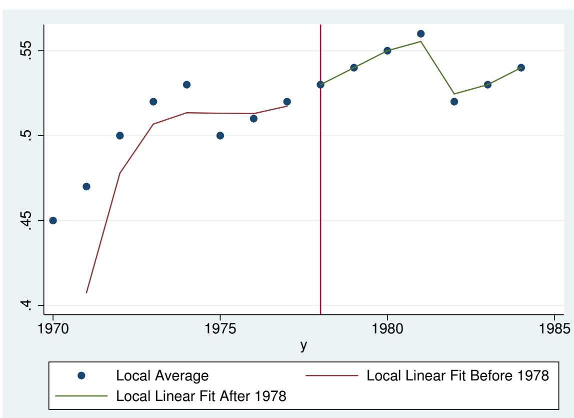
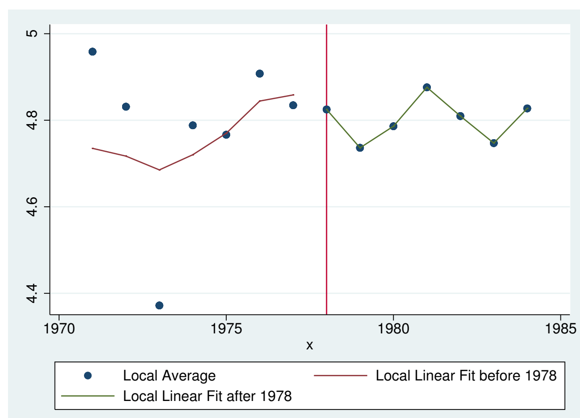



Figure 1.8: Male labor market outcome by birth-cohorts




(a) Male Labor Force Participation




(b) Log Male Real Wage

## 1.5 Results

The outcome variables for this analysis are highest education obtained, labor force participation and log earnings. Table 1.2 presents the results from the D3 and D4 specifications (equation (1.1) and equation (1.2)). I include age, household size, a dummy for rural areas and landownership of households as covariates. Row 1 is from the D3 specification while Row 2 is from the D4 specification. The third and fourth rows are from the D4 specification where young refers to the young cohort (age 25 to 27 years) and middle (28 to 32 years) refers to the middle cohort after collapsing the data by birth-cohorts, districts, gender and survey years. 1995 HIES and 2010 HIES are used. Standard errors are clustered by age and districts. The average treatment effect of the program is estimated to be an increase of about 0.21 years of education in the treated areas, net of males of the same cohort and females of the older cohort (age more than 32 years) in the D3 specification and an increase of 0.31 years of education, net of males of same age, females of older cohort and females of same age in non-treated areas in the D4 framework. Among the females who received the stipend for only a fraction of their secondary education, the increase was about 0.35 years of education and among those who got it for all five years, the increase was about 0.52 years of education. Since the data for Row 3 and Row 4 are collapsed, I cannot compare the effects on the young and the middle cohort with that from Row 2.

Labor force participation is estimated to be 2 percentage points higher among the program beneficiaries using the triple differences framework and about 3 percentage points higher using the D4 specification. The estimated increase in labor force participation among women who received the stipend for only a fraction of their secondary education is 2 percentage points while it is 7 percentage points among women who received it for all five years of their secondary school. The average treatment effect among the intended program participants is an increase of 35 percent in real earnings, net of males and females of older cohorts in the treated areas. In case of the D4 approach, the average impact is an increase of 48 percent in earnings among the intended program participants. The effects on earnings seem to be larger for the middle cohort than for the younger cohort. This might be due to the effects of diminishing marginal returns as I find later that the effects on wages decline as the number of years of exposure increases.

Table 1.2: Impact Estimates using Difference-in-Difference-in-Difference (D3) and Difference-in-Difference-in-Difference-in-Difference (D4) Framework

| | 1 | 2 | 3 |
|---|---|---|---|
| | Education Attained | Labor Force Participation | Log Real Earnings |
| $S_i$*T*f | 0.21 | 0.02 | 0.35 |
| | [0.10]* | [0.01]* | [0.14]** |
| $S_i$*p*T*f | 0.31 | 0.03 | 0.48 |
| | [0.18]** | [0.01]*** | [0.24]** |
| young*p*T*f | 0.52 | 0.07 | 0.42 |
| | [0.26]* | [0.01]*** | [0.30] |
| middle*p*T*f | 0.35 | 0.02 | 0.57 |
| | [0.19]* | [0.01]* | [0.26]** |

Notes: *p≤0.1, **p≤0.05, ***p≤0.01. Standard errors in brackets. Details of the regression given in A.4. The data used are from 1995 and 2010 HIES. Row 1 is the result from the D3 regression and Row 2 is the result from the D4 regression. Row 3 and 4 are from the D4 regression after collapsing the data by birth-cohort, gender and districts. Each regression includes age, household size, land ownership and rural location as covariates. Row 3 and Row 4 gives the estimate for young and middle cohort in a D4 framework respectively. Young cohort (row 3) are those who received a stipend for 5 years while the middle cohort (row 4) received it for less than 5 years.

Table 1.3 shows the regression discontinuity results from the local linear regression while Table 1.4 shows the results from the quartic polynomial fit (equation 1.3). For all the regression discontinuity results, I use the sample of 25 to 50 year old females from the 1995 to 2010 quinquennial Household Income and Expenditure Surveys. I collapse my dataset into cell means by birth-year, gender, district and survey year and weight by population. For Table 1.3 and Table 1.4, I restrict the sample to females in treated areas. The fit from the local linear regression predicts an increase of 0.6 years of completed education, a 9 percentage points increase in labor force participation and a 55 percent increase in wages between the 1978 and 1979 birth-cohorts as shown in Table 1.3.

Table 1.3: Estimated Effect of FSSAP on Highest Education Attained, Labor Force Participation and Log Real Earnings using Local Linear Regression

| 1 | 2 | 3 |
|---|---|---|
| Education Attained | Labor Force Participation | Log Real Earnings |
| 0.575 | 0.087 | 0.551 |
| [0.001]*** | [0.003]*** | [0.131]*** |

Notes: *p≤0.1, **p≤0.05, ***p≤0.01. Standard error in brackets. The data are from 1995, 2000, 2005 and 2010 HIES. It is collapsed by birth-year, gender, districts and survey years and then restricted to only females in treated areas. Local linear regression uses a rectangular kernel and optimal bandwidths.

The parametric regression discontinuity specification (Table 1.4) includes controls for a birth-cohorts quartic polynomial and an indicator for whether the cohort received free secondary education for some fraction of their secondary school. The polynomial fit predicts an increase in educational attainment between the 1978 and 1979 cohorts of about 1 year with a standard error of 0.34 and an $R^2$ of 0.90 which is shown in Column 1 of Table 1.4. The estimate does not vary much if I include controls for age as shown in Column 2 of Table 1.4. This specification predicts an increase of 6.4 percentage points in the labor force participation rate with a standard error of 0.03 and an $R^2$ of 0.89 and a 72 percent increase in real earnings with a standard error of 0.278 and an $R^2$ of 34 percent. Columns 2, 3, 5, and 6 also include age controls as a quartic polynomial. All regressions are weighted by cell size and the standard errors are clustered by birth-cohorts and districts using Huber-White standard errors. The results do not change substantially when I include the age-quartic controls.

Table 1.4: Estimated Effect of FSSAP on Highest Education Attained, Labor Force Participation and Log Real Earnings using Polynomial Fit

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| | Education Attained | | Labor Participation | | Log Real Earnings | |
| | 1.01 | 1.00 | 0.06 | 0.07 | 0.73 | 0.98 |
| | [0.34]*** | [0.35]*** | [0.03]** | [0.03]** | [0.28]** | [0.35]*** |
| Birth polynomial | Quartic | Quartic | Quartic | Quartic | Quartic | Quartic |
| Age polynomiall | None | Quartic | None | Quartic | None | Quartic |

Notes: *p≤0.1, **p≤0.05, ***p≤0.01. Standard errors given in brackets. The data are from 1995, 2000, 2005 and 2010 HIES. It is collapsed by birth-year, gender, districts and survey years and then restricted to only females in treated areas. The regressions are estimated by a polynomial fit which includes the quartic polynomial in birth-year and an indicator for the cohorts born in or after 1978.

Table 1.5 shows results from equation (1.4) and equation (1.5). The outcome variables are educational attainment and labor force participation. Columns 1, 2, 4 and 5 are results from equation (1.4) while columns 3 and 6 are from equation 1.5. Columns 2, 4, 5 and 6 include illiteracy rate in the districts, the fraction living in rural areas in the district and the fraction working in manufacturing as the covariates. All regressions include time fixed effects and a quartic polynomial in birth year. The results suggest that in the treated areas, the intended program participants have about a year more education and are 5 percentage points more likely to be in the labor force. However, including the non-treated areas in the analysis decrease the estimates significantly. I find that the intended program participants have only 0.21 year of more education and are 2 percentage points more likely to be in the labor force.

Table 1.5: Effects of FSSAP on grade attainment and labor force participation using RD

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| | Years of Education | | | Labor Force Participation | | |
| $X_c$ | 0.19 | 0.19 | 0.10 | -0.02 | -0.02 | -0.02 |
| | (0.90) | (0.89) | (0.46) | (-1.04) | (-1.07) | (-0.97) |
| f | -1.93** | -1.93** | -1.59** | -0.85** | -0.85** | -0.83** |
| | (-20.94) | (-21.15) | (-11.05) | (-172.80) | (-175.07) | (-113.19) |
| $X_c$*f | 0.95** | 0.94** | 0.69** | 0.05** | 0.05** | 0.02 |
| | (4.78) | (4.86) | (3.50) | (4.06) | (4.21) | (1.35) |
| Illiteracy rate | | -7.21** | -6.90** | | 0.06** | 0.05** |
| | | (-15.97) | (-17.23) | | (2.63) | (2.10) |
| Rural | | -0.42 | -0.57** | | -0.05** | -0.03** |
| | | (-1.09) | (-2.12) | | (-2.56) | (-2.18) |
| Manufacturing | | -9.56** | -5.28** | | 0.70** | 0.67** |
| | | (-4.64) | (-2.59) | | (4.69) | (4.96) |
| p | | | 0.35** | | | 0.02** |
| | | | (2.31) | | | (2.74) |
| F*p | | | -0.34** | | | -0.02** |
| | | | (-2.08) | | | (-2.53) |
| $X_c$*f*p | | | 0.21** | | | 0.02** |
| | | | (2.05) | | | (2.56) |
| Time dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Birth Polynomials | Quartic | Quartic | Quartic | Quartic | Quartic | Quartic |
| Regional dummies | Yes | No | No | Yes | No | No |
| Observations | 6297 | 6349 | 7783 | 6298 | 6298 | 7732 |
| $R^2$ | 0.218 | 0.200 | 0.208 | 0.878 | 0.872 | 0.873 |

Notes: *p≤0.1, **p≤0.05. *t* statistics in parentheses.The data are from 1995, 2000, 2005 and 2010 HIES. It is collapsed by birth-year, gender, districts and survey years. Column 1, 2, 4 and 5 are the results from equation 1.4 and Column 3 and 6 are from equation 1.5. Column 1 and 4 include district level fixed effects while Column 2, 3, 5 and 6 include district level covariates.

Table 1.6 presents the results from equations (1.4) and (1.5) for log earnings. In treated areas, we see an increase in earnings of about 37 percent among the intended program participants. Including the non-treated areas, the increase in earnings decline to 5 percent and it is statistically insignificant.

Table 1.6: Effects of FSSAP on log real earnings using RD

| | 1 | 2 | 3 |
|---|---|---|---|
| | Log Real Earning | | |
| $X_c$ | -0.04 | -0.04 | -0.07 |
| | (-0.52) | (-0.52) | (-1.02) |
| f | -0.08 | -0.10 | -0.35** |
| | (-1.01) | (-1.19) | (-4.00) |
| $X_c$*f | 0.37** | 0.37** | 0.33** |
| | (2.84) | (2.71) | (2.88) |
| Illiteracy rate | | -0.96** | -1.21** |
| | | (-5.78) | (-8.55) |
| Rural | | -0.15 | -0.51** |
| | | (-0.94) | (-4.54) |
| Manufacturing | | 3.98** | 3.62** |
| | | (3.26) | (3.98) |
| p | | | -0.05 |
| | | | (-1.05) |
| f*p | | | 0.24** |
| | | | (3.09) |
| $X_c$*f*p | | | 0.05 |
| | | | (0.35) |
| Time dummies | Yes | Yes | Yes |
| Birth Polynomials | Quartic | Quartic | Quartic |
| Regional dummies | Yes | No | No |
| Observations | 2671 | 2671 | 3380 |
| $R^2$ | 0.519 | 0.530 | 0.536 |

Notes: $t$ statistics in parentheses. Same specification as Table 1.5

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$

Table 1.7 reports the estimates from the first stage and reduced form regressions after differentiating between young and middle cohorts in the treated areas. I also use this specification for all the other results given below. FSSAP can be associated with an increase in education of 0.83 years for the young cohort of female and of 0.67 years for the middle birth-cohorts. It contributed to an increase in labor force participation by almost 5 percentage points for both the young and middle cohort of female. Earnings increased by 27 percent for the middle cohort. For the young cohort, the effect on earnings is positive but statistically insignificant. Including regional controls instead of dummies has little effect on the results.

Table 1.7: Effects of FSSAP on grade attainment, labor force participation and log real earnings with two birth-cohorts and using RD

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| | Grade attainment | | Labor force participation | | Log earnings | |
| young | 0.569 | 0.56 | -0.052 | -0.051 | 0.243 | 0.243 |
| | [0.163]*** | [0.163]*** | [0.011]*** | [0.011]*** | [0.071]*** | [0.071]*** |
| middle | 0.130 | 0.127 | -0.044 | -0.043 | 0.109 | 0.108 |
| | [0.197] | [0.197] | [0.009]*** | [0.009]*** | [0.055]* | [0.055]* |
| young *f | 0.832 | 0.834 | 0.047 | 0.047 | 0.148 | 0.146 |
| | [0.230]*** | [0.230]*** | [0.009]*** | [0.009]*** | [0.175] | [0.176] |
| middle * f | 0.672 | 0.674 | 0.046 | 0.046 | 0.269 | 0.271 |
| | [0.222]*** | [0.222]*** | [0.009]*** | [0.009]*** | [0.089]*** | [0.092]*** |
| Female | -1.963 | -1.965 | -0.05 | -0.05 | -0.547 | 0.545 |
| | [0.127]*** | [0.127]*** | [0.008]*** | [0.006]*** | [0.053]*** | [0.053]*** |
| Rural | | -2.331 | | 0.086 | | -0.688 |
| | | [1.55] | | [0.141] | | [0.469] |
| Manufacturing | | -2.580 | | -0.700 | | 3.159 |
| | | [4.904]*** | | [0.325]** | | [2.029] |
| Illiteracy rate | | -1.211 | | 0.06 | | -1.11 |
| | | [0.43]*** | | [0.01]*** | | [0.23]*** |
| Birth Polynomials | Quartic | Quartic | Quartic | Quartic | Quartic | |
| Time dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Regional dummies | Yes | No | Yes | No | Yes | No |

Notes: *p≤0.1, **p≤0.05, ***p≤0.01. Standard errors given in brackets. The sample is restricted to treated areas. Middle refers to cohort born between 1978 and 1982 and young refers to cohort born after 1982.

Next I try to investigate which sector is attracting the more educated female labor force. Table 1.8 shows that the policy is helping women to find jobs in the formal sector. Females with secondary education are more likely to work in formal institutions that pay them more. The likelihood that a FSSAP beneficiary, who got the stipend for the full 5 years, found a job in the formal sector is 3 percentage points higher than for males in the same age range. Similarly, those who got the stipend for less than 5 years are also 8 percentage points more likely to find a job in the formal sector than are men in the same cohort. They are also more likely to work for the government and less likely to work in manual agricultural jobs than men in the same cohort. Table 1.9 shows that the policy provides little incentive for females to stay beyond the secondary school and I find no policy effect on labor force participation and earnings for women with more than a secondary school education. SSC (Secondary School Certificate) in Table 1.9 refers to the nationwide board exams that all students sit for after completion of grade 10.

Table 1.8: Effects of FSSAP on likelihood of working in agriculture, non-agriculture, informal sector and formal sector with two birth-cohorts and using RD

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| | Government | Private Agriculture | Rural | Informal sector | Formal sector |
| young*f | 0.04 | -0.16 | 0.01 | -0.143 | 0.03 |
| | [0.005]*** | [0.02]*** | [0.02] | [0.025]*** | [0.017]* |
| middle*f | 0.04 | -0.19 | 0.005 | -0.02 | 0.088 |
| | [0.006]*** | [0.03]*** | [0.02] | [0.018] | [0.016]*** |

Notes: *p≤0.1, **p≤0.05, ***p≤0.01. Standard errors given in brackets. The sample is restricted to treated areas. Middle refers to cohort born between 1978 and 1982 and young refers to cohort born after 1982. Government includes working for central government, local government and public corporations. Private agriculture includes those working on own or somebody elseÕs land. Formal sector employment consists of regular workers while informal sector employment include day laborers and self-employed workers.

Table 1.9: Estimated Effect of FSSAP on Highest Education Attained, Labor Force Participation and Log Real Earnings with Different Samples

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| | Grade attainment | | Labor force participation | | Log earnings | |
| | Below SSC | After SSC | Below SSC | After SSC | Below SSC | After SSC |
| young*f | 0.568 | 0.01 | 0.045 | 0.026 | 0.149 | -0.016 |
| | [0.174]*** | [0.084] | [0.01]*** | [0.015]* | [0.097] | [0.293] |
| middle*f | 0.688 | 0.016 | 0.046 | 0.027 | 0.186 | -0.288 |
| | [0.153]*** | [0.054] | [0.01]*** | [0.022] | [0.07]** | [0.184] |

Notes: *p≤0.1, **p≤0.05, ***p≤0.01. Standard errors given in brackets. The sample is restricted to treated areas. Middle refers to cohort born between 1978 and 1982 and young refers to cohort born after 1982.

Migration between the districts of Bangladesh is less than 1% giving the opportunity to identify the policy effect through differences in exposure by gender, location and age. My results do not vary much by the different methods of identification that I have implemented in this paper adding support to my conclusion that FSSAP not only increased women′s educational levels but also led to better labor market outcomes for women.

### 1.5.1 Specification Tests

According to Imbens and Lemieux (2007), there are two concerns that generally arise in RD designs. The first is the possibility of other changes in covariates at the same cutoff value and the second is the sensitivity to bandwidths. During the second half of the 1900s, Bangladesh went through boom in the garment sector and rapid rural to urban migration. I find evidence that neither experienced a jump at the cutoff point considered in the above analysis. The coefficients of the dependent variables from local linear regressions do not vary much with different bandwidths either. I also test whether impacts vary with the years of exposure to the program.

#### Test involving covariates

Figure 1.9 shows that there is no discontinuity in regional controls for the 1978 birth-cohorts, who were the first graduates to get the secondary stipend. The fraction employed in manufacturing in the district decreased for the birth-cohorts of 1978 although it has an upward trend overall. Similarly, the fraction living in rural areas decreased but there is no discontinuity around 1978.

#### Sensitivity to bandwidth choice

In the regression discontinuity analysis, I use the optimal bandwidth suggested by Imbens and Lemieux (2007) (bandwidth is 10 for the baseline specification). If the estimates and standard errors are critically dependent on bandwidth choices, then the results may be suspicious. I re-estimate the local linear regressions with bandwidths of 5 and 20 and find that although the estimates and standard errors vary, the variation is small. The overall results still hold providing credibility to the answers above. Table 1.10 provides the results with different bandwidths.

Figure 1.9: Covariates

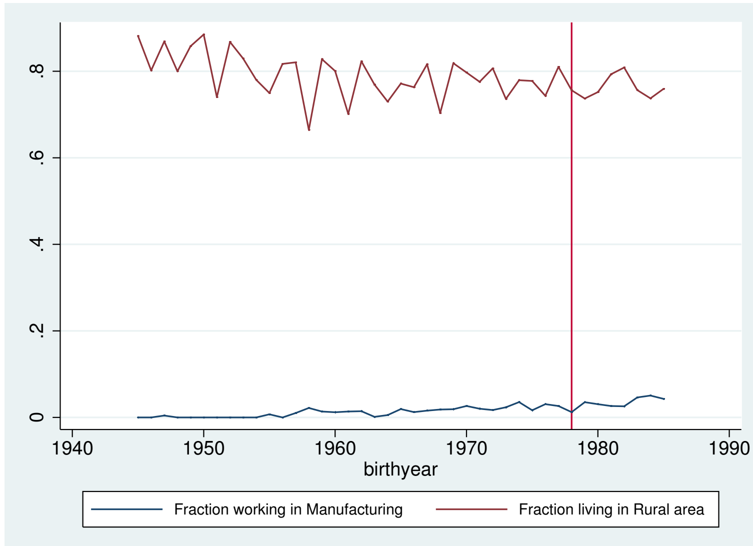



Table 1.10: Bandwidth

| | 1 | 2 | 3 |
|---|---|---|---|
| Bandwidth | Grade attainment | Labor force participation | Log earning |
| 10 | 0.57 | 0.087 | 0.635 |
| | [0.001]*** | [0.003]*** | [0.044]*** |
| 5 | 0.462 | 0.082 | 0.529 |
| | [0.000]*** | [0.000]*** | [0.000]*** |
| 20 | 0.397 | 0.08 | 0.699 |
| | [0.021]*** | [0.007]*** | [0.059]*** |

Notes: *p≤0.1, **p≤0.05, ***p≤0.01. Standard errors given in brackets. Table 1.3 is replicated with different bandwidths.

### Other Sensitivity Tests

Next, I study how the number of years of exposure to the program affects educational attainment, labor force participation and earnings. All the birth-cohorts before 1978 got zero years of exposure, while the 1978 birth-cohorts got 1 year, the 1979 birth-cohorts got 2 years, the 1980 birth-cohorts got 3 years, the 1981 birth-cohorts got 4 years and the

1982 birth-cohorts and above got 5 years of FSSAP. Table 1.11 shows that one year of the subsidy increases education by 0.18 years, labor force participation by 1 percentage point and earnings by 4%.

Table 1.11: Years of exposure to the FSSAP

| 1 | 2 | 3 |
|---|---|---|
| Grade attainment | Labor force participation | Log earning |
| 0.183 | 0.011 | 0.039 |
| [0.041]*** | [0.002]*** | [0.02]* |

Notes: *p≤0.1, **p≤0.05, ***p≤0.01. Standard errors given in brackets.

Using dummies for each of the above birth-cohorts, I find that grade attainment and labor force participation increased almost uniformly for each of the birth-cohorts due to FSSAP. Earnings increased for the first three cohorts but the fourth and fifth years of subsidy seem to have no significant effect on earnings. This might be due to diminishing returns of the policy as the years of exposure increase. The effects on the 1976 and the 1977 birth-cohorts are insignificant suggesting that grade repetition and delayed school entry did not lead to older cohorts benefiting from the program. Table 1.12 provides the details.

Table 1.12: Effect on different birth-cohorts

| | 1<br>Grade attainment | 2<br>Labor force participation | 3<br>Log earning |
|---|---|---|---|
| 1976 | 0.457 | 0.017 | -0.154 |
| | [0.499] | [0.050] | [0.119] |
| 1977 | -0.08 | 0.033 | 0.159 |
| | [0.098] | [0.047] | [0.106] |
| 1978 | 1.012 | 0.044 | 0.403 |
| | [0.127]*** | [0.009]*** | [0.055]*** |
| 1979 | 1.049 | 0.051 | 0.444 |
| | [0.127]*** | [0.009]*** | [0.053]*** |
| 1980 | 0.422 | 0.048 | 0.311 |
| | [0.128]*** | [0.009]*** | [0.052]*** |
| 1981 | 0.843 | 0.043 | 0.02 |
| | [0.183]*** | [0.009]*** | [0.05] |
| 1982 and above | 0.407 | 0.046 | 0.165 |
| | [0.194]** | [0.006]*** | [0.126] |

Notes: *p≤0.1, **p≤0.05, ***p≤0.01. Standard errors given in brackets. Specification same as Table 1.7.

## 1.6 Conclusion

Conditional cash transfer (CCT) programs are rapidly becoming a key policy instrument used by developing countries to reduce poverty and increase human capital investment. Unlike middle-income countries, developing countries have implemented gender targeted or location specific CCTs instead of the mean-tested CCTs. FSSAP is one such program with the aim of promoting education and economic empowerment among women.

This paper studies the effects of FSSAP on the labor markets of Bangladesh. It extends the present literature by studying the long-term effects of the world's first conditional cash transfers. Since most social protection programs, implemented by the international development organizations, are now conditional cash transfers, it is important to look at the long-term effects. The results suggest that FSSAP not only increased school attainment for girls but also the likelihood of them being in the labor force and earnings more than both their older and male counterparts. It also helped them to find a job in the formal sector. The program can be associated with an increase in education level completed by at least 0.21

years, labor force participation by 2 percentage points and earnings by 33 percent. Although past research has found significant positive effect on enrollment, questions have arisen about whether the stipend program has any real effect on school completion and female empowerment. This paper is an attempt to answer some of these questions.

The results suggest that the government may have been correct to expand the programs to boys since 2009. Returns to education in developing countries are higher than in developed countries and providing assistance to credit constrained individuals to help them to go to schools has both short and long-term beneficial impacts on the economy (Carneiro and Heckman, 2002). However, since the governments of the developing countries have limited resources, further research may study how to target the policy to those who have the highest returns and the effect of that on the labor market.

# Chapter 2

# Testing a Roy Model with Multiple Markets: The Case of International Migrants from Bangladesh

## 2.1 Introduction

According to the Migration and Remittances Factbook (World Bank, 2011), Bangladesh is one of the top ten emigration and remittance receiving countries with 5.4 million of estimated emigrants as of 2010 and remittances of $11.1 billion in 2010. While there have been a few descriptive studies (Mahmood (1993) and Mahmud (1989)) done on the determinants of remittances and their impact on the Bangladesh economy, to the best of my knowledge, there has not been any empirical research investigating the reasons behind such a large labor migration out of Bangladesh. This paper studies the role that earnings play in shaping individual migration decisions. This paper investigates two questions. First, does the destination, which offers the highest earnings relative to Bangladesh, attract the most migrants? Second, is the relative stock of more-educated migrants in a destination increasing in the earnings difference between high and low-skilled workers?

Table 2.1: Destinations

| | Percent |
|---|---|
| Saudi Arabia | 35.42 |
| Kuwait | 6.29 |
| UAE | 24.65 |
| Bahrain | 1.78 |
| Singapore | 3.63 |
| Malaysia | 14.19 |
| Gulf | 7.34 |
| Asia | 2.96 |
| West | 3.04 |
| Africa | 0.69 |

Table 2.1 shows the destination choices of Bangladeshi migrants. About 75% of the migrants choose the Middle East region as their destination of which 35% go to Saudi Arabia (KSA) and 24% go to the UAE (United Arab Emirates). About 20% of the migrants choose an Asian country while about 3% choose a country from Europe or the Americas (West) with the majority choosing the USA or UK. Table 2.2 shows that migrants with secondary school or lower are more likely to migrate to the gulf countries while migrants with a college degree are more likely to go to the western countries. Average monthly income of the migrants is highest in the western countries, where 20% of the migrants have a college degree and lowest in Malaysia, where 82% of the migrants have only a primary education. Even within the gulf region, we see that income in Saudi Arabia, which receives 35% of the migrants from Bangladesh, is higher than in the UAE, which receives about 24% of the migrants. Saudi Arabia also gets more migrants with secondary school and college education than the UAE. This leads to the question of whether the difference in earnings among different education groups in different countries plays a role in the migration flow patterns of Bangladeshis.

Table 2.2: Summary Statistics

| Destination | Income | Illiterate | Primary | Secondary | College | Married |
|---|---|---|---|---|---|---|
| Bangladesh | 2.92 | 0.44 | 0.44 | 0.08 | 0.03 | 0.86 |
| | ( 4.59) | ( 0.50) | ( 0.50) | (0.27) | (0.17) | (0.35) |
| Saudi Arabia | 18.98 | 0.10 | 0.79 | 0.10 | 0.01 | 0.68 |
| | (18.57) | (0.30) | (0.41) | (0.30) | (0.11) | (0.47) |
| Kuwait | 20.30 | 0.10 | 0.78 | 0.11 | 0.01 | 0.64 |
| | (19.59) | (0.29) | ( 0.41) | (0.31) | ( 0.11) | (0.48) |
| UAE | 17.21 | 0.09 | 0.82 | 0.08 | 0.01 | 0.55 |
| | (14.48) | (0.29) | (0.39) | (0.27) | (0.09) | (0.50) |
| Bahrain | 18.21 | 0.14 | 0.79 | 0.06 | 0.01 | 0.60 |
| | (8.53) | (0.35) | (0.41) | (0.24) | (0.11) | (0.49) |
| Singapore | 31.34 | 0.02 | 0.72 | 0.23 | 0.03 | 0.40 |
| | (24.50) | (0.14) | (0.45) | (0.42) | (0.17) | (0.49) |
| Malaysia | 15.79 | 0.09 | 0.82 | 0.09 | 0.01 | 0.57 |
| | (13.23) | (0.29) | (0.39) | (0.28) | (0.08) | (0.50) |
| Gulf | 20.72 | 0.12 | 0.79 | 0.09 | 0.01 | 0.67 |
| | (22.47) | (0.32) | (0.41) | (0.28) | (0.09) | (0.47) |
| Asia | 29.90 | 0.14 | 0.68 | 0.13 | 0.05 | 0.60 |
| | (45.71) | (0.35) | (0.47) | (0.33) | (0.23) | (0.49) |
| West | 83.69 | 0.01 | 0.32 | 0.47 | 0.20 | 0.67 |
| | (75.68) | (0.07) | (0.47) | (0.50) | (0.40) | (0.47) |
| Africa | 33.30 | 0.02 | 0.72 | 0.22 | 0.04 | 0.51 |
| | (50.16) | (0.15) | (0.45) | (0.41) | (0.21) | (0.50) |

Mean is reported and standard error is given in parentheses. Income is monthly income in thousands of Taka.

Table A.5 shows that 98% of the migrants are men. 78% of the migrants had employment contracts before entering the destination country as Table A.6 presents. On average it takes 5.87 months for a migrant to secure a contract and process the visa. Table A.6 also shows that the West is the most expensive destination with an average migration cost of Tk. 604,000 (USD 7723). Among the gulf countries, Bahrain is the most expensive with an average migration cost of Tk. 259,000 (USD 3313) and Kuwait is the least expensive with an average cost of Tk. 191,000 (USD 2442).

In the same spirit as Dahl (2002), I estimate a multi-market Roy (1951) model of mobility and earnings with self-selection where different destination countries are modeled as having different earnings and benefits for workers with different schooling levels. The Roy model predicts that an increase in reward to skill in a destination should cause immigration from the source country to rise and the mix of migrants to become more skilled. I first estimate the selection probabilities semi parametrically by grouping individuals with the same discrete characteristics together and taking cell means for the probabilities associated with different migration paths. In the second step, I use these migration probabilities in the correction function for each country to get consistent estimates of the potential individual earnings at each location. In the final step, I test the responsiveness of migration flows to differences in earnings and amenities across countries and also estimate a discrete location choice model of migration.

The rest of the paper is structured as follows. Section 2 gives the literature review. Section 3 presents a simple model of migration. Section 4 describes the data and Section 5 the estimation approach. Section 6 discusses the results and Section 7 concludes.

### 2.2 Literature Review

There has been a substantial literature focusing on the role of income on bilateral migration flows. Bertoli et al. (2012), Clemens et al. (2008) and McKenzie and Rapport (2010) use individual level data while other studies use aggregate data (Grogger and Hanson (2011), Belot and Hatton (2013), Ortega and Peri (2009), and several others) to study the determinants of international migration. Bertoli et al. (2012) is closest to the study of this paper but they do not look into how migration flows respond to earnings differences across education levels and they focus on migration due to the economic crisis of 1990 in Ecuador. This paper extends the literature by empirically looking into migration from a developing country, Bangladesh, by education level using individual level data for all destination countries.

There have been a few descriptive analyses of labor migration from Bangladesh. Mahmud (1989) explains the overall macro impact of migration on the Bangladesh economy. Mahmood (1993) claims that the possible motive for migration is the prospect of a higher wage and better job opportunities. Mahmood (1993), Rahman (2011b) and Rahman (2011a) are, however, skeptical about whether migration is economically advantageous given the migration cost and exploitation by recruiting agencies. Due to data limitation on agencies, I cannot explicitly model the role of agencies but I am able to investigate whether migration flows respond to migration cost in this paper.

This paper uses the approach developed in Dahl (2002) who analyzed internal migration in the USA. Dahl's approach addresses the problem that failing to account for unobserved factors that simultaneously affect the earnings and the migration decision can bias the estimates of a multi-choice model.

## 2.3 Model

The Roy (1951) model I consider is similar to the one used by Dahl (2002) with multiple markets, choice based on utility maximization, and an unspecified distribution of latent skills. One difference is that I consider only one country of birth and nine destination countries or "residence" countries rather than fifty-one birth and residence states as considered by Dahl (2002). All individuals start out in Bangladesh, k=0. Each individual decides whether to migrate to one of the eight potential destinations (k = 1, 2, 3, 4, 5, 6, 7, 8). Each individual compares the expected utility from migrating to each destination to the expected utility of staying in the home country and then opts for the utility maximizing alternative. Following Dahl (2002), I assume that utility consists of an additively separable function of log earnings, person-specific and destination-specific taste factor. The taste factor represents the non-wage determinants of migration, such as the cost of moving, amenity differences between destinations and the psychological costs and benefits associated with moving to a different country. That is,

$$U_{ik} = y_{ik} + t_{ik}, \quad k = 0, ..., 8 \tag{2.1}$$

where $U_{ik}$ indexes utility, $y_{ik}$ is log earnings and $t_{ik}$ is a term for an individual's taste for moving from country 0 to k. The deviation of an individual's earnings and tastes if they were to work in country k from their expected values conditional on observables are

$$y_{ik} - E[y_{ik}|z_i] = u_{ik}, \quad k = 0, ..., 8 \tag{2.2}$$

$$t_{ik} - E[t_{ik}|z_i] = w_{ik}, \quad k = 0, ..., 8 \tag{2.3}$$

where $z_i$ is a vector of individual characteristics and $w_{ik}$ and $u_{ik}$ are the error terms. The utility for individual $i$, associated with the destination k is the sum of a deterministic component $V_{ik}$ and an unobserved stochastic component $e_{ik}$, where $V_{ik}$ $=E[y_{ik}|z_i]$+$E[t_{ik}|z_i]$ and $e_{ik}$ $=w_{ik}$+$u_{ik}$.

$$U_{ik} = V_{ik} + e_{ik}, \quad k = 0, ..., 8 \tag{2.4}$$

If $M_{ik}$ is the indicator for whether the individual migrates to destination k, the selection equation is

$$\begin{aligned} M_{ik} &= 1 \text{ iff } V_{ik} + e_{ik} \geq V_{im} + e_{im} \quad \text{for all m} \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{2.5}$$

However, an individual's earnings are not observed for all countries, but only for the destination country that is his utility maximizing option. An ordinary least squares earnings regression will give biased estimates of the returns to education and of the other coefficients since

$$E[u_{ik}|y_{ik} \text{ is observed}] = E[u_{ik}|V_{ik} + e_{ik} \geq V_{im} + e_{im}] \neq 0 \tag{2.6}$$

The direction and size of the bias depends on the joint distribution of $u_{ik}$ and the error terms from the migration equations, $e_{i0}$-$e_{i1}$,...,$e_{i0}$-$e_{i8}$. With a two sector model, the usual approach specifies the joint distribution of the error terms in the outcome equation and the single selection equation to be bivariate normal (Heckman, 1979). For polychotomous choice models, most previous work makes distributional assumptions which greatly simplify the form of the selectivity bias (McFadden, 1984). In order to avoid imposing restrictive assumption on the selection bias, I follow Dahl (2002) and express the selectivity bias as a function of the probability of selection given covariates. The bias correction in the earnings equation ideally should include the individual's potential migration probabilities to all destinations. For this analysis, I define the correction function for destination k as an unknown function of $p_{ik}$, such that k≠0, which is the individual's propensity to migrate to destination k and $p_{i0}$ which the individual's propensity to stay back in Bangladesh. The earnings equations can then be written as a multi-index, partially linear models which depends on a subset of migration probabilities.

$$y_{ik} = \alpha_k + \alpha_k z_i + M_{ik} * \psi(p_{ik}, p_{i0}) + \epsilon_{ik} \tag{2.7}$$

Supposing that individuals behave according to comparative advantage, an equation describing the migration flow to country k for those with education level h in terms of earnings and amenities is

$$\ln(p_k^h) = \theta_0^h + \theta_1(y_k^h - y_0^h) + \theta_2^h(A_k - A_0) + \theta_3^h D_k + v_k^h \tag{2.8}$$

where $y_k^h$ is the average earnings of individuals with education level h in destination k, $A_k$ is a vector of amenities associated with destination k and $D_k$ is the cost of moving to k. $y_k^h$ and $y_0^h$ are both expected earnings and need to be estimated. A similar equation for education group f can also be defined. Schooling level does not change the package of amenities offered by the state. However, the value that individuals in different education classes place on those amenities is expected to differ. With destination-specific earnings, it is assumed that individuals belonging to any education class respond identically to a given difference in earnings. Differencing the log migration flows of the two education groups then yields

$$ln(p_k^h) - ln(p_k^f) = (\theta_0^h - \theta_0^f) + \theta_1((y_k^h - y_k^f) - (y_0^h - y_0^f)) + (\theta_2^h - \theta_2^f)(A_k - A_0) + (\theta_3^h - \theta_3^f)D_k + (v_k^h - v_k^f) \tag{2.9}$$

To estimate the migration decision model (equation (2.5)), most researchers have either used conditional logit or nested logit model. While the conditional logit model requires the assumption of “independence of irrelevant alternatives", the nested logit model depends on the researcher's decision about which choices belong to which nest and which choices are independent. To compare the results, I estimate the migration choice model by both conditional logit and nested logit model. For the nested logit model, I assume that the unobserved component of the migration equation ($e_{ik}$) is correlated across destinations, k = 1 to 8, for a given individual. There is no correlation between ($e_{i0}$) and the analogous terms for the other locations, k = 1 to 8.

## 2.4 Data

The analysis uses individual level data on non-migrants in Bangladesh and on migrants to the destination countries. I combine two datasets for this purpose - the "Household Income and Expenditure Survey (HIES), 2009-2010" and the "Bangladesh Household Remittance Survey (BHRS) of 2009". They are both nationally representative surveys and after merging I weight them appropriately.

The data on the non-migrants come from the HIES while that on migrants come from the BHRS. The HIES has a sample size of 27,572 non-migrants while the BHRS has a sample size of 12,893 migrants. I restrict the sample to men as migrants from Bangladesh are predominantly male. I also restrict the sample to those who are aged between 15 and 65 years and to those who are employed. I group the destination countries that do not receive a substantial number of migrants from Bangladesh by their region. For example, Gulf includes the countries that belong to the Middle East region but do not get many Bangladeshi migrants, for example Qatar, Oman, Iraq and Iran. Asia includes countries that belong to Asia and Australia for example Japan, China, Hong Kong, Australia and New Zealand. West includes countries like the USA, UK, Italy, and Germany. Africa includes countries like Egypt and Nigeria. A full list of the countries can be found in the appendix (Table 2.3). Due to the small sample size of migration to Africa, I exclude Africa from the estimation.

The merged dataset contains information on age, marital status, household size, education, years since migration, employment status, occupation, occupation before migration and labor earnings. We use labor earnings at the time of the survey as a proxy for expected earnings in each of the nine countries. After restricting the dataset, the final sample contains 9,899 migrants and 12,242 non-migrants or stayers.

The amenity variables include cost of migration, distance, average temperature, five- year average unemployment and population density. The temperature data come from the World Bank's Climate data (API), unemployment and population density data are from the World Development indicator, the data on distance are from Geonames and the migration stock data are from BMET.

### 2.5 Estimation Approach

The estimation is done in three steps. First, I estimate the probability that an individual follows a given migration path. Second, I estimate the selection corrected earnings of each individual at each location. Third, I test the appropriateness of Roy model of migration and earnings by estimating how migration flows respond to differences in corrected earnings and other amenities and estimate the migration choice equation.

#### 2.5.1 Calculating Migration Probabilities

The first step to correct for self-selection involves estimating the migration probabilities for individuals. Dahl (2002) assumes that people with the same level of education who are the same age and similar in relevant characteristics are affected by the differences in earnings, moving costs, state taxes and other amenities in the same way. Following his approach, I build different cell structures for Bangladesh and for the eight destination countries, to take into account the different sample size in the locations. For Bangladesh, I define 32 different cells defined by four education levels (illiterate, primary, secondary and college), age group (less than or equal to 35 years and more than 35 years), marital status and household size (less than or equal to 4 and more than 4). The average cell size is 777 stayers with a minimum of 14 and maximum of 1349. Since there are fewer observations for migrants, their groupings are coarser. I assign individuals to groups by education level and by whether their relatives helped them to migrate in 8 different cells[1]. The average cell size for migrants is 3076 to any of the eight destinations with a minimum of 70 and maximum of 4330. The fraction of individuals in a cell who migrate to any of the destinations estimates the probability that any individual in the cell will follow the same migration path. The probability that an individual i chooses location k is calculated as the proportion of individuals with observable characteristics in the same cell that choose to live and work in location k, i.e. the $\{\hat{p_{ik}}\}$. Table A.8 in the appendix summarizes the overall variability in the cell migration probabilities for stayers and movers by education class for all 9 countries. Looking at stayers, we see that illiterate and college educated Bangladeshis are more likely to remain in Bangladesh. Looking at movers, we see that individuals with primary and secondary education are more likely to migrate out of Bangladesh (Table A.9).

### 2.5.2 Selection Corrected Estimates of the Return to Education and Earnings

Dahl (2002) employs a second order polynomial series expansion for estimating the correction function. I estimate the following equation for log earnings:

$$y_{ik} = \alpha_k + \theta_k z_i + f_k(\hat{p_{ik}}, \hat{p_{i0}}) + \epsilon_{ik} \tag{2.10}$$

where $z_i$ includes a constant, a college-graduate dummy, a secondary school dummy, a primary school dummy, age and its square, household size, a marital status dummy and five occupational dummies. The function $f_k(\hat{p_{ik}}, \hat{p_{i0}})$ is Dahl's correction polynomial. The exact form is a second order polynomial in the retention probability for stayers ($\hat{p_{i0}}$), and a second

[1] Using the same cell division for migrants and stayers does not affect the main results.

order polynomial in the probability for migration to country k ($\hat{p_{ik}}$), plus an interaction term. The standard errors of the coefficient estimates of equation (2.10) need to take into account that the cell probabilities are estimated. To correct for the extra sampling variability arising from the imputed migration probabilities, I report bootstrapped standard errors. I also estimate the asymptotically correct standard error (Murphy and Topel, 1985)[2] but due to the relatively small number of observations, the Murphy and Topel covariance matrix does not differ much from the standard covariance matrix. Next I predict the log earnings in all nine locations for all individuals in the sample using the above estimated log earnings equations.

### 2.5.3 Testing the Roy Model and Migration Choice

Next, I estimate equation (2.9) to test whether migration flows respond to earnings differences between the secondary school educated and the primary school educated. I aggregate the log earnings and the migration probabilities by education levels. The equation is estimated using OLS regression but the standard errors are corrected for heteroskedasticity and the extra sample variability due to using estimated log earnings. I also estimate a discrete choice migration model that allows for correlation across alternatives. The deterministic part of utility can be written as:

$$V_{ik} = \alpha \hat{y_{ik}} + x_i' \beta_k \tag{2.11}$$

The main explanatory variable for the probability of choosing a particular location is the estimate of log earnings at that location. This estimate takes into account that migrants

[2]The Murphy and Topel estimator of the asymptotically correct covariance matrix is

$$(\hat{x'}\hat{x})^{-1}\hat{x'}\hat{\Gamma}\hat{v}(\hat{p})\hat{\Gamma}\hat{x}(\hat{x}'\hat{x})^{-1} + \hat{\sigma}^2(\hat{x}'\hat{x})^{-1}$$

where $\hat{x}$ denotes the matrix of explanatory variables appearing in the wage equations. The first term accounts for the sampling variability of the estimated probabilities and the second term is the usual covariance matrix. $\hat{\Gamma}$ and $\hat{v}(\hat{p})$ are block diagonal matrices. Let $n_c$ denote the sample size of cell "c". Each block of $\hat{\Gamma}$ is an $n_c$x2 matrix containing the derivatives of the correction functions with respect to the two migration probabilities evaluated at their estimated values. These derivatives are easily calculated for polynomial expansions. Each block of $\hat{v}(\hat{p})$ is an estimate of the 2x2 covariance matrix for the estimated mover and stayer probabilities for a cell.

may not be a random sample of the original population, allowing for the degree of self-selection to vary by destination. I also include controls for all the variables that were used in the previous step. $x_i$ is the individual-specific characteristics and following the normal convention, $\beta_1$ is normalized to zero. $x_i'\beta_k$ is the net gain from migration that can vary at the individual level, reflecting differences in gender, age, education, marital status and household composition. I assume that $e_{ik}$, the stochastic error term, has a Generalized Extreme Value distribution, which gives rise to the nested logit model. There are two nests. One nest is a singleton containing only Bangladesh. The second nest contains the other eight destination countries.

## 2.6 Results

This section presents the estimation results. I estimate one earnings equation for each of the nine destination countries. The dependent variable is the log of monthly earnings in Bangladeshi Taka. The right hand side includes age, age squared, dummies for primary, secondary and college degree, household size, a dummy for marital status and five occupational dummies. The occupational dummies represent laborer, operator, salesman, businessman and other professions (including doctors, engineers, teachers). For each country, I estimate a Mincer regression (labeled Uncorrected), which does not control for self-selection and a model (labeled Corrected) that controls for self-selection by including the polynomial with the cell probabilities. Table 2.3 reports the Mincer estimates while Table 2.4 reports the Dahl estimates for all the nine countries. Both tables support the idea that there is a wage premium for the college educated over the secondary school educated and for the secondary school educated over the primary school educated. The size of the wage premium does not vary substantially between the two models. Overall returns to education are high in Asia and low in the Middle East.

Table 2.3: Mincer Equations

| | BD | KSA | Kuwait | UAE | Singapore | Malaysia | Gulf | Asia | West |
|---|---|---|---|---|---|---|---|---|---|
| Age | 0.05*** | 0.02** | 0.02 | 0.01 | -0.02 | 0.03*** | -0.02 | 0.07* | 0.03 |
| | (17.17) | (3.27) | (1.17) | (1.34) | (-0.56) | (3.44) | (-1.39) | (2.31) | (1.18) |
| $Age^2$ | -0.00*** | -0.00** | -0.00 | -0.00 | 0.00 | -0.00** | 0.00 | -0.00 | -0.00 |
| | (-15.84) | (-2.60) | (-0.78) | (-0.50) | (0.59) | (-3.12) | (1.54) | (-1.46) | (-0.61) |
| Primary | 0.22*** | 0.11*** | 0.12* | 0.14*** | 0.23* | 0.09*** | 0.07* | 0.42*** | 0.04 |
| | (11.40) | (6.20) | (2.37) | (7.60) | (2.50) | (3.42) | (2.20) | (4.17) | (0.15) |
| Secondary | 0.67*** | 0.35*** | 0.49*** | 0.35*** | 0.24* | 0.25*** | 0.29* | 1.10*** | 0.35 |
| | (19.65) | (10.39) | (5.33) | (8.49) | (2.06) | (4.77) | (2.39) | (5.68) | (1.32) |
| College | 1.06*** | 0.52*** | 0.83* | 0.36** | 0.41 | 0.44 | 0.64 | 1.96*** | 0.30 |
| | (24.97) | (4.66) | (2.34) | (2.88) | (1.83) | (1.07) | (1.78) | (6.15) | (1.21) |
| Household | 0.03*** | -0.00 | -0.01 | -0.00 | -0.00 | 0.00 | 0.00 | 0.03 | -0.01 |
| | (9.14) | (-0.79) | (-1.03) | (-0.29) | (-0.23) | (0.47) | (0.62) | (1.86) | (-1.21) |
| Married | 0.24*** | -0.01 | -0.02 | 0.03 | 0.09 | 0.00 | 0.15** | -0.32** | 0.18 |
| | (12.31) | (-0.25) | (-0.32) | (1.09) | (1.15) | (0.03) | (3.13) | (-2.77) | (1.94) |
| Laborer | 1.00*** | 0.23*** | 0.14 | 0.11** | -0.04 | -0.03 | 0.19** | -0.09 | 1.41*** |
| | (17.23) | (7.57) | (1.83) | (2.94) | (-0.29) | (-0.87) | (2.95) | (-0.51) | (3.54) |
| Operator | 1.15*** | 0.37*** | 0.41*** | 0.27*** | -0.02 | 0.11** | 0.38*** | 0.04 | 1.31** |
| | (18.89) | (12.18) | (7.67) | (8.49) | (-0.17) | (3.24) | (5.87) | (0.20) | (3.02) |
| Sales | 1.17*** | 0.32*** | 0.18* | 0.38*** | -0.27 | 0.08 | 0.38*** | -0.04 | 1.39*** |
| | (18.99) | (9.53) | (2.11) | (6.45) | (-0.81) | (1.39) | (3.55) | (-0.19) | (3.39) |
| Business | 0.94*** | 1.06*** | 0.98*** | 0.84*** | 1.19* | 0.37 | 1.39*** | -0.20 | 1.48*** |
| | (5.18) | (14.65) | (4.72) | (4.80) | (2.37) | (1.48) | (6.28) | (-0.63) | (3.52) |
| Professional | 1.30*** | 0.30*** | 0.22** | 0.38*** | 0.06 | 0.09 | 0.43*** | 0.42 | 1.52*** |
| | (21.12) | (7.99) | (3.23) | (7.28) | (0.40) | (1.74) | (5.44) | (1.71) | (3.69) |
| Constant | -1.17*** | 1.93*** | 2.00*** | 2.10*** | 3.35*** | 1.90*** | 2.60*** | 0.73 | 1.65* |
| | (-14.73) | (15.86) | (6.08) | (16.40) | (6.14) | (11.84) | (10.99) | (1.27) | (2.56) |
| Observations | 12242 | 3354 | 579 | 2325 | 296 | 1341 | 757 | 278 | 247 |
| $R^2$ | 0.302 | 0.146 | 0.197 | 0.129 | 0.134 | 0.067 | 0.203 | 0.462 | 0.213 |

$t$ statistics in parentheses, p-value in square brackets

$^*$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 2.4: Wage Equations with Dahl's corrections

| | BD | KSA | Kuwait | UAE | Singapore | Malaysia | Gulf | Asia | West |
|---|---|---|---|---|---|---|---|---|---|
| Age | 0.04*** | 0.02** | 0.03 | 0.01 | -0.02 | 0.03* | -0.02 | 0.08* | 0.02 |
| | (13.39) | (3.28) | (1.39) | (1.26) | (-0.65) | (2.51) | (-1.45) | (2.14) | (0.46) |
| $Age^2$ | -0.00*** | -0.00* | -0.00 | -0.00 | 0.00 | -0.00* | 0.00 | -0.00 | -0.00 |
| | (-12.25) | (-2.42) | (-0.85) | (-0.48) | (0.68) | (-2.18) | (1.81) | (-1.35) | (-0.00) |
| Primary | 0.20*** | 0.11*** | 0.11** | 0.13*** | 0.23** | 0.09*** | 0.06 | 0.41*** | 0.02 |
| | (11.32) | (5.35) | (2.62) | (8.56) | (2.67) | (4.33) | (1.53) | (4.04) | (0.07) |
| Secondary | 0.67*** | 0.34*** | 0.54*** | 0.34*** | 0.31* | 0.22*** | 0.31*** | 1.04*** | 0.45 |
| | (20.19) | (11.57) | (6.81) | (6.58) | (2.52) | (4.39) | (3.80) | (5.38) | (1.55) |
| College | 1.07*** | 0.53*** | 0.96** | 0.36* | 0.59 | 0.44 | 0.63 | 1.88*** | 0.50 |
| | (21.41) | (3.42) | (2.93) | (2.34) | (1.76) | (1.19) | (1.76) | (5.23) | (1.62) |
| Household | 0.03*** | -0.00 | -0.01 | -0.00 | -0.01 | 0.00 | 0.00 | 0.03* | -0.02 |
| | (9.94) | (-0.85) | (-1.07) | (-0.32) | (-0.43) | (0.21) | (0.69) | (1.98) | (-1.20) |
| Married | 0.25*** | 0.01 | -0.05 | 0.03 | 0.11 | 0.01 | 0.15* | -0.32** | 0.22 |
| | (10.58) | (0.53) | (-0.96) | (1.21) | (1.26) | (0.32) | (2.37) | (-2.66) | (1.81) |
| Laborer | 0.99*** | 0.22*** | 0.13 | 0.12** | -0.04 | -0.03 | 0.19*** | -0.10 | 1.36** |
| | (15.47) | (9.00) | (1.73) | (2.73) | (-0.28) | (-0.74) | (3.86) | (-0.39) | (2.98) |
| Operator | 1.15*** | 0.36*** | 0.39*** | 0.28*** | -0.01 | 0.11** | 0.38*** | 0.03 | 1.28** |
| | (18.29) | (14.59) | (5.76) | (8.04) | (-0.09) | (2.71) | (9.65) | (0.11) | (2.80) |
| Sales | 0.31*** | 0.16 | 0.39*** | -0.30 | 0.07 | 0.37*** | -0.05 | 1.35** | 1.18*** |
| | (9.62) | (1.65) | (5.65) | (-0.77) | (1.42) | (3.44) | (-0.19) | (2.95) | (17.39) |
| Business | 0.95*** | 1.03*** | 0.96*** | 0.83*** | 1.24* | 0.37 | 1.39*** | -0.22 | 1.46** |
| | (7.04) | (12.06) | (4.62) | (6.86) | (2.24) | (1.40) | (5.46) | (-0.65) | (3.01) |
| Professional | 1.31*** | 0.29*** | 0.21** | 0.39*** | 0.09 | 0.09 | 0.43*** | 0.43 | 1.49** |
| | (17.83) | (8.91) | (2.79) | (9.74) | (0.65) | (1.56) | (4.78) | (1.45) | (3.22) |
| Constant | -1.00*** | 1.98*** | 2.04*** | 2.15*** | 3.34*** | 2.09*** | 2.67*** | 0.61 | 2.12** |
| | (-10.86) | (11.92) | (6.31) | (15.35) | (6.57) | (9.58) | (10.64) | (0.90) | (2.66) |
| Wald Test for $f_k$ | 8.88** | 27.4*** | 8.37* | 10.5* | 8.17 | 14.3** | 7.15 | 3.0 | 5.51 |
| | [0.01] | [0.00] | [0.08] | [0.06] | [0.09] | [0.01] | [0.21] | [0.70] | [0.24] |
| Observations | 12242 | 3354 | 579 | 2325 | 296 | 1341 | 757 | 278 | 247 |
| $R^2$ | 0.303 | 0.150 | 0.207 | 0.133 | 0.155 | 0.078 | 0.208 | 0.466 | 0.233 |

$t$ statistics in parentheses, p-value in square brackets

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Following Dahl (2002) I test the impact of the correction terms by using Wald test statistics[3]. To conduct the test I use the asymptotically correct covariance matrix. Table 2.4 shows that all the correction terms except for Gulf, Asia and the West, are significant at the 10% significance level. However, testing for the statistical significance of the differences between the corrected and uncorrected coefficients of secondary and college education by using a Hausman test shows that the difference is not statistically significant (Table A.10 and A.11). Only for the secondary school return for Singapore, do we find the difference between corrected and uncorrected estimates to be statistically significant at the 5% level. Overall, correcting for self-selection does not seem to make a large quantitative difference to the return estimates and the predictions of earnings. I find that for Bangladesh, the corrected return to primary school is lower than the uncorrected return suggesting that stayers in Bangladesh with primary school degree have systematically higher wage draws than those who decide to leave.

Table 2.5 gives the estimation results for equation (2.9) using earnings differential and amenity variables. It compares the migration flows of secondary school to primary school graduates. Since most migrants from Bangladesh are either secondary school or primary school graduates, I compare these two groups[4]. All the specifications indicate that the relative stock of more-educated migrants in a destination increases with the earnings differences between skilled and unskilled workers. The first two columns of Table 2.5 show that migration responds to both uncorrected and corrected earnings difference between the two education groups. To understand the impact these different differentials have on migration patterns, I consider the coefficient estimate from column 2 of Table 2.5. The relative migration flow between those with secondary versus primary school education is 1.83% more to the country that offers 1% more in relative earnings, net of the relative earnings in Bangladesh. Columns

[3]The test that Dahl(2002) conducts is

$$\eta[\hat{v}(\hat{p})]^{-1}\eta \sim \chi^2 \text{with df} = rank(\eta)$$

where $\eta$ is the vector of coefficients for the terms in the correction function and $\hat{v}(\hat{p})$ is the appropriate block from the covariance matrix of the estimated wage equation.

3 and 4 of Table 2.5 include the amenity variables in the regression. Migrants with secondary school education are less likely to migrate to a country that is warmer than Bangladesh and where the total migration stock in 2000 is lower.

Table 2.5: Responsiveness of secondary school to primary school migration flows to differences in earnings and amenities

| Dependent variable: $\ln(p_k^h)$ - $\ln(p_k^f)$ | | | | | |
|---|---|---|---|---|---|
| | Uncorrected | Corrected | Corrected with amenities | | |
| Δ Earnings | 1.92*** | 1.83*** | 2.24*** | 1.92*** | 1.47*** |
| | (49.78) | (44.23) | (39.11) | (34.82) | (26.54) |
| Distance | | | 0.054 | 0.29*** | ?0.008 |
| | | | (0.15) | (14.94) | (-0.29) |
| ΔUnemployment | | | 0.021*** | 0.11*** | 0.11*** |
| | | | (4.77) | (17.38) | (18.61) |
| ΔPopulation density | | | 0.00017*** | 0.00012*** | 0.0008*** |
| | | | (37.39) | (23.94) | (16.28) |
| ΔTemperature | | | -0.16*** | -0.33*** | -0.26*** |
| | | | (-30.99) | (-36.37) | (-26.36) |
| Cost | | | -0.0000056*** | -0.000022*** | -0.000006*** |
| | | | (-11.75) | (-16.40) | (-3.7) |
| ΔAge Dependency Ratio | | | | -0.031*** | -0.0168*** |
| | | | | (-23.42) | (-10.28) |
| Migration Stock in 2000 | | | | | -0.17 *** |
| | | | | | (16.32) |
| Constant | 0.29*** | 0.26*** | 0.30*** | 0.32*** | 0.31*** |
| | (43.92) | (36.66) | (124.80) | (138.67) | (139.7) |
| $R^2$ | 0.33 | 0.31 | 0.64 | 0.67 | 0.69 |

$t$ statistics in parentheses, h is secondary school and f is primary school
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

To estimate the discrete location choice model, the deterministic part of the utility associated with each location is assumed to be a function of log earnings, a country-specific intercept, and individual characteristics like age, education dummies, marital status and household size.

[4]Table A.12 compares migration flows between college and secondary school graduates.

The country specific intercepts account for the differences in policies, institutions, culture, and cost of living expenses between the destination countries. The vector of individual controls is meant to capture individual differences in the cost of migration. Table 2.6 presents my estimation of the discrete choice model using nested logit while Table A.13 presents the result from a conditional logit model that assumes that there is no correlation across alternatives in unobservable. I also include location-specific amenity variables in the nested logit model and the results are given in Table A.14. In all the specifications, the coefficient on earnings is positive and statistically significant, suggesting that higher expected earnings in a particular country increase the probability to locate there. Education variables are also significant in all the models suggesting that conditional on expected earnings, education still plays a role in shaping international migration decisions. Conditional on expected earnings, having a college degree is associated with a lower probability of migrating to the Middle East and a higher probability of migrating to Asia. Similarly, having a college or a secondary degree is associated with a higher probability of migrating to Europe. The Likelihood Ratio test statistic of the restriction, that there is no correlation across alternatives in unobservable (dissimilarity parameter ($\tau$) =1), leads to strong rejection of the conditional logit model in favor of the nested logit model. The additive random utility model restricts $0<\tau<1$. In my model, the estimated value of $\tau$ is 101.54 that means that although the model is mathematically correct, the fitted model is not consistent with the additive random utility model.

Table 2.6: Nested Logit without Amenities

| | KSA | Kuwait | UAE | Singapore | Malaysia | Gulf | Asia | Europe |
|---|---|---|---|---|---|---|---|---|
| Corrected wage | 1.72*** | | | | | | | |
| | (24.43) | | | | | | | |
| | KSA | Kuwait | UAE | Singapore | Malaysia | Gulf | Asia | Europe |
| Secondary | 12 | 13.91 | -29.32* | 211.26* | -40.99* | -20.83 | -52.97* | 362.45* |
| | (1.66) | (0.90) | (-2.20) | (2.67) | (-2.29) | (-1.33) | (-1.88) | (2.8) |
| College | -15.15 | -0.53 | -76.09* | 180.35* | -109.68 | -80.55 | 10.36 | 446.71* |
| | (-0.99) | (-0.01) | (-2.25) | (2.62) | (-1.86) | (-1.43) | (0.29) | (2.74) |
| Log Likelihood: = -26434.218 | | | | | | | | |
| Dissimilarity Coefficient: 101.54 | | | | | | | | |
| LR test for IIA ($\tau = 1$): $\chi^2(1) = 457.86$ Prob> $\chi^2 = 0.0000$ | | | | | | | | |

$t$ statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

## 2.7 Conclusion

This paper tests whether migrants from Bangladesh choose the destination countries that offer them the highest earnings. Using individual level data from Bangladesh, I estimate the role of earnings as a determinant of international migration flows, accounting for self-selection as proposed by Dahl (2002). The main finding is that international migration flows respond not only to earnings differences between countries but also to earnings differences between different education groups. Conditional on expected earnings, the result indicates that education is an important determinant of migration. Overall the results support the view that labor migration is an income maximizing option for Bangladeshis and policies to promote international migration should also address the development of human capital.

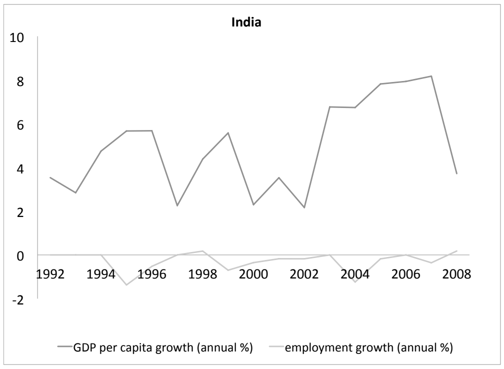
# Chapter 3

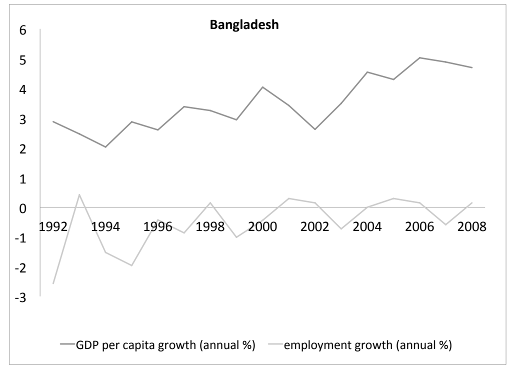
# Labor Market Effects of Employment Guarantee Program in India

## 3.1 Introduction

South Asian countries, namely Bangladesh and India have been undergoing rapid economic growth averaging around 6 to 8 percent since 2002 but employment growth has been slow and almost nonexistent (Figure 3.1). In India, formal sector employment has declined since 2002. Recent study by the World Bank (2010) finds that there has been a deceleration in the growth of wages and that a large fraction of workers are trapped in jobs with low earnings, poor working conditions and no unemployment insurance in both of these countries. Labor is the principal asset for both of these countries and in order to ensure pro-poor and sustainable growth, these countries are finding it challenging to create productive jobs for workers. Bino (2008) finds that inequality in India has been on the rise - while the richest 1 percent of the population saw a 71 percent increase in real income between 1987 and 1999, 20% of the population is still living on less than a dollar. Sharma (2006) and Mazumder (1983) have argued that government policies and other labor market institutions may have been responsible for the inflexibility in the labor market, especially the persistent large share of rural sector in the labor force and the small share of the formal sector.

Figure 3.1: *Employment growth and GDP per capita growth in Bangladesh and India*




The first initiative taken by the governments of most South Asian countries to address concerns over lack of permanent job availability and social security is to provide protection through special employment programs for the unskilled workers. The initiative taken by the Indian government is the National Rural Employment Guarantee Act (MNREGA). Similar policies have also been taken by governments of Bangladesh and Sri Lanka, namely '100 days of employment generation program for the poor' and 'short term employment generation program' respectively. MNREGA is a job guarantee scheme enacted on August 2005 and implemented in February 2006 with the twin objectives of rural development and employment generation. It provides a legal guarantee for 100 days of employment in every financial year to any rural household willing to do unskilled work at the statutory wage of Rs. 60 ($1.1) daily in 2006 and was later increased to Rs. 120 ($2.2) in 2009. If employment is not provided within the first 15 days of application, then an unemployment allowance is provided. This wage under MNREGA is substantially higher than both the average wage of Rs. 48 ($0.88) that the rural informal workers were making and than the average wage of Rs. 52 ($0.96) that the urban informal workers were making in 2004-2005. The work involved is mostly manual in nature including work like dam and road construction and irrigation. The project provided 1435.9 million person days of employment in 1.78 million projects in

2007 and covered almost 99% of the districts by 2008. The outlay for the project was 4 billion dollars in 2008 and it was increased to 8.92 billion dollars in 2010. Theoretically this employment guarantee program should affect the labor market in various ways. It provides an extra source of income for rural unskilled people which means that they now will have less incentive to migrate to urban areas avoiding congestion and overcrowding. The projects include infrastructure development that may induce productivity growth in rural areas. However, if the program is financed by taxing the already small formal sector firms, it might be detrimental to the formal sector which is concentrated in the urban areas.

In this paper, I build a three-sector model of the Indian labor market with two types of workers, skilled and unskilled, and study the impacts of such targeted policy on the labor market outcomes, welfare and inequality in India. The model that I build incorporates a rural sector, where only unskilled workers are present, an urban informal sector where both skilled and unskilled workers are present and the formal sector where only skilled workers can work. I use the data from 2004-2005 to calibrate the model to the pre-MNREGA economy. I then ask what would be the aggregate effects on the overall labor market if MNREGA were increasing the rural wage by a certain amount? Since the model in this analysis is a steady state model, the aggregate effects it predicts are long-run effects. Nonetheless, I look at data from 2009 to see what aggregate labor market effects have occurred so far. I find that although MNREGA has increased welfare and employment for the unskilled workers in the rural sector, it has adverse effects on the skilled workers decreasing both formal employment and welfare.

The remainder of the paper is organized as follows - Section 2 gives a brief summary of MNREGA and Section 3 and 4 discusses the literature review and some stylized facts about the Indian labor market. Section 5 describes the model and characterizes the steady-state equilibrium. Section 6 reports the calibration procedure and Section 7 discusses the policy implications. Section 8 reports the empirical results and Section 9 concludes.

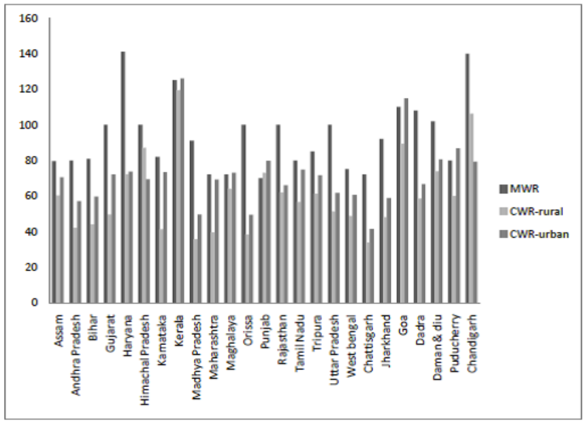
## 3.2 MNREGA- A Brief Overview

According to "Gazette of India," MNREGA offers one hundred days of employment to any rural household willing to do unskilled manual work for which they will receive Rs. 60 or the minimum wage, whichever is higher. It covered 200 districts in February 2006, 320 districts in April 2007 and 625 districts in April 2008 out of the 640 districts in India. To apply for a job under MNREGA, the household must apply for a job card to the Gram Panchayat (village leaders) with the names, ages and addresses of all the working members in the family. All registered households are then entitled for 100 days of employment in a given financial year. The work site under the program has to be within 5 km radius of the applicant's village. If no such work is possible, they are paid 10% more to meet the additional travel and living expenses. It is a nationwide program covering all districts except Chennai, Kolkata, Hyderabad, New Delhi, and Mumbai.

The central government provides the funds for the wages and 75 percent of the material costs. No machinery is used. The state government pays for 25 percent of the material cost. If the person is not provided work within 15 days of job application, then he is eligible for an unemployment allowance for each day after the 15 days until the state finds work for him. The unemployment allowance cannot be less than 25 percent of the minimum wage rate in the first 30 days of the financial year and it must be 75 percent of the minimum wage rate in the remaining period of the financial year. The state government has to bear the cost of the unemployment allowance.

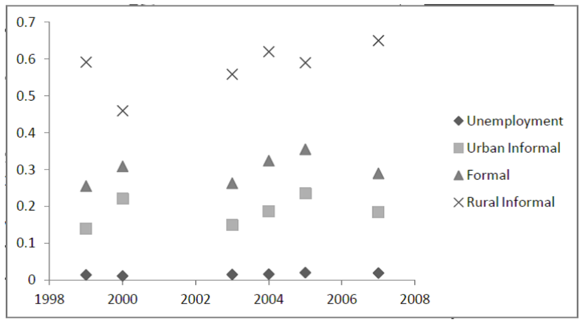
The projects under the scheme should be for rural development. The works include water conservation, water harvesting, drought proofing, irrigation, land development, flood control, rural connectivity and creation of durable assets. The manual nature of the work theoretically should only attract unskilled laborer and as such no other eligibility criteria is imposed other than being a rural household. Farrington and Sjoblom (2008) argue that MNREGA has the potential to stimulate rural growth through the establishment of productive physical

Figure 3.2: *Daily wage rate of casual laborers working in rural areas (CWR-rural), in urban areas (CWR-urban) in 2004-2005 (source: NSSO surveys) and the wage rate (MWR) under MNREGA (source: Gazette of India).*




assets and the influx of funds through wages. The program that preceded MNREGA, the Maharashtra employment program increased agricultural wages by making laborers reluctant to accept anything less than the official minimum wage. MNREGA can stabilize income for rural households, as more work will be available in the agricultural off-period. It may discourage workers from migrating to the urban areas in search of employment. Figure 3.2 shows that in most of the states, the statutory wage rate under MNREGA was higher than the rate casual workers were receiving in both rural and urban areas, providing evidence for higher income under MNREGA. Figure 3.3 shows that after the implementation of MNREGA, the rural labor force increased while urban informal and formal sector employment decreased supporting the hypothesis of lower migration to urban areas.

Figure 3.3: *Trends in labor composition according to location and sector of employment. (source: NSSO surveys)*




## 3.3 Literature Review

A few papers have studied the effects of MNREGA on the rural labor markets. Imbert and Papp (2012), using a difference-in-difference approach, find that it increases employment by 0.3 days and private sector casual wages by 4.5 percent. Azam (2012), using the same method, finds that the public sector employment increased by 2.5 percent while wages increased by 8 percent. However, Zimmerman (2012) finds no effect on employment and only a small increase in private sector wages for women using regression discontinuity method. All these papers study the direct effect of MNREGA on the rural labor market. The size and the scope of the program suggest that it should have indirect effects on the urban labor market. This paper extends the literature by looking at the potential equilibrium spillover effects of MNREGA on the urban segment of the Indian labor market.

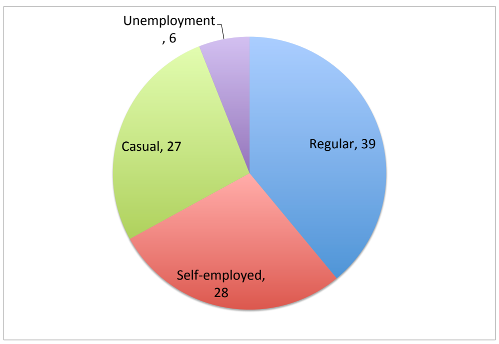
A substantial literature has developed that studies segmented labor markets in developing countries and implications of different policies. These papers are more suited to middle income countries and are focused mostly on Latin American countries. To the best of my knowledge, this paper is the first attempt to model labor markets of developing countries like India where dynamism in the informal sector plays an important role in transmission of any shock to the labor market.

The first paper to study a segmented labor market is by Harris and Todaro (1970). Their main result, the Todaro Paradox, states that creating urban jobs may increase rather than decrease urban unemployment because of the increase in rural-urban migration which may outweigh the positive effects of creating jobs. Fields (1975) extends the Harris - Todaro model to include job-search behavior. Zenou (2008) models the formal sector as the one with search frictions while the informal sector is perfectly competitive. Workers are homogenous and they find that reducing unemployment benefits or the firms$'$ entry cost in the formal sector induces job creation and formal employment, reduces the size of the informal sector but has ambiguous effect on wages. Satchi and Temple (2009) differs from Zenou (2008) in having no unemployment state and examines how growth affects the size and the outcomes of the informal and formal sectors. Albrecht et al. (2009) extends the literature by studying the impacts of labor market policy on informality in an economy with heterogeneous workers and sector specific production technologies. They also find that severance taxes greatly increase average employment duration in the formal sector, reduces unemployment and the number of formal sector workers, and a payroll tax reduces average employment duration in the formal sector and the number of formal sector workers. Gutierrez et al. (2009) adds a perfectly competitive agricultural sector to Albrecht et al. (2009) to study the effects of the financial crisis on labor markets in developing countries.

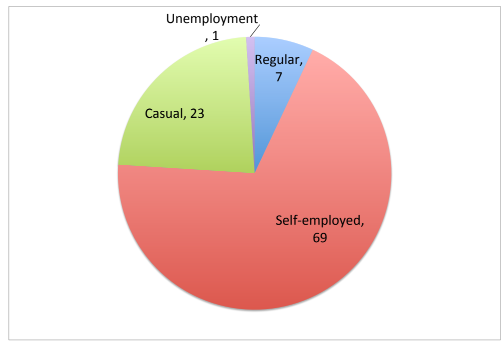
### 3.4 Indian Labor Market

World Bank (2010) reports that the Indian labor market is divided into two sectors: the formal sector, which is concentrated in the urban regions and the informal sector, which is concentrated in the rural regions. The rural sector employs 59 percent of the labor force while the urban sector employs 41 percent. Dividing the labor market according to location, Figure 3.4 reveals that 92% of the rural workers work for the informal sector in 2004. In the urban areas, 39 percent of the labor force work as regular workers in the formal sector and 55 percent work for the informal sector. Unemployment has been very low at around 1% in the rural areas which suggests that MNREGA acts more like a wage subsidy than employment guarantee since most people in the rural households are already employed although in low waged temporary jobs.

Figure 3.4: *Labor composition according to location in 2004*




(a) Urban



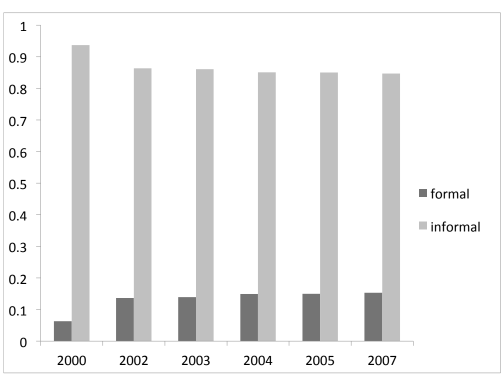

(b) Rural

World Bank (2010) and Bino (2008) both report that more than 80% of the workforce is in the informal sector (Figure 3.5). In theory, the informal sector is distinguished from the formal sector by the absence of legal protection and formal recognition by the government. Informal sector firms tend to be small (employing less than 10 people for India), paying no tax to the government and with no capability of workers to organize themselves in unions and without formal employment contracts.

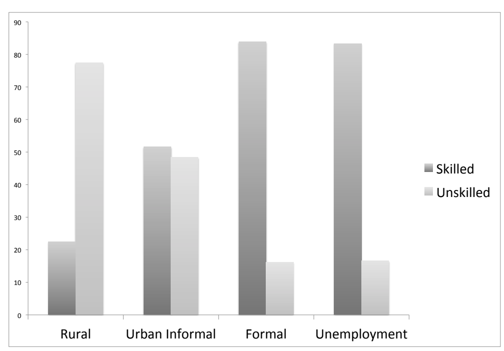
The National Commission for Enterprises in the Unorganized Sector (NCEUS) (Bino, 2008) reports that 98% of the agricultural sector, 70% of the industrial sector and 72% of the service sector are informal. The non-agricultural informal work includes manufacturing, construction, trade and repair, transport storage and other services. The agricultural informal workers include the agricultural laborers. NCEUS points out that the profile of workers in the informal sector reveals that lack of education is a major factor, which prevents the informal sector workers from improving their position in the labor market. Further, NCEUS indicates that lower levels of education among the informal workers create vulnerability at two levels. Firstly, it denies access to good jobs in the formal sector, and secondly it confines the workers to mostly manual jobs. The situation in rural areas is worse, as illiteracy coupled with a poor asset base and landlessness force individuals to work as casual laborers in the agricultural sector (Sen, 1966).

The formal sector in India is small (less than 20%) and according to definition used by Bino (2008) includes the regular worker category in the NSSO (National Sample Survey Organization) surveys. Bino (2008) reports that the share of regular employment is mainly concentrated in three economic activities namely public administration, manufacturing, and education. Regular workers also require a minimum level of schooling that is essential to work for the formal sector. They also find that unemployment rates are higher in urban areas than in rural areas. The unemployment rate is seen to increase with educational attainment and is particularly high among those with a secondary level of education and higher (Figure 3.5).

Figure 3.5: *Labor composition according to skill and sectors*




(a) Sector




(b) Education

## 3.5 Model

I model India as a small open economy with three sectors- formal sector, urban informal sector and rural informal sector respectively. The formal sector produces an intermediate good using only skilled labor while the urban informal sector uses both skilled and unskilled labor to produce another intermediate good. These two intermediate goods are sold in a competitive market and transformed into a final consumption good. The price of the final good is normalized to 1. The production technology of the final good is a Cobb-Douglas, $Y = Y(F, I)$. F is the formal sector output and I is the urban informal sector output. The production function is such that $Y_F > 0, Y_{FF} < 0, Y_I > 0$ and $Y_{II} < 0$, $Y_{FI} > 0$ and $Y_{IF} > 0$. The prices are then $P_f = Y_F$ and $P_I = Y_I$. In line with Chakrabarti and Kundu (2009), my calibration exercise finds that the formal sector good and the urban informal sector good are complements. The production function of the formal sector is such that one skilled worker produces $A_f$ units of output. The production function of the urban informal sector is a CES production function, $I = I(l, h)$, where $l$ is the number of unskilled workers, $h$ is the number of skilled workers, $I_l > 0$, $I_h > 0$, $I_{hh} < 0$, $I_{ll} < 0$, $I_{hl} > 0$, $I_{hl} > 0$ and the Inada conditions are satisfied. Estimation results in Table A.15 reveal that the skilled and the unskilled workers are complements in the urban informal sector. The urban informal sector is perfectly competitive. The wage each worker receives is his/her marginal productivity. $w_l = I_l$ and $w_h = I_h$ where $w_l$ is the wage of urban informal sector unskilled worker and $w_h$ is the wage of urban informal sector skilled worker. The urban formal sector has only skilled workers and is characterized by search frictions. In the formal sector, firms and workers come together by a matching function that depends on the number of people looking for formal sector jobs and the number of vacancies, v, available. The skilled worker in the formal sector gets wage, $w_f$, but at exogenous rate $\delta$ he can lose the job. The value of a formal sector job for the skilled workers is given by

$$rF = w_f + \delta \max[U - F, H - F]$$

where F is the value of formal sector employment, U is the value of unemployment and H is the value of urban informal sector employment for the skilled workers. When the skilled worker loses the formal sector job, he can either move to the informal sector or be unemployed and look for other formal sector jobs. If unemployed, he has a better opportunity to find a job in the formal sector than if he was working in the urban informal sector since he has more time to look for a job. The unemployed gets some unemployment benefits, b, and then with rate m($\theta$) finds a job in the formal sector. The matching rate of the unemployed is

$$m(\theta) = \frac{M(v, o + \lambda s_I)}{o + \lambda s_I}$$

where $\theta = \frac{v}{u+\lambda h}$, o is the number unemployed and $s_I$ is the number of skilled workers employed in the urban informal sector, $m'(\theta) > 0$, $\frac{d(m(\theta)/\theta)}{d\theta} < 0, \lim m(\theta) \rightarrow 0$ as $\theta \rightarrow \infty$, and $\lim m(\theta) \rightarrow \infty$ as $\theta \rightarrow 0$. $\lambda \in (0, 1)$ is an exogenous search intensity parameter for skilled workers employed in the urban informal sector. The value of unemployment is

$$rU = b + m(\theta)[F - U]$$

and the value of an urban informal sector job for the skilled workers is

$$rH = w_h + \lambda m(\theta)[F - H]$$

In equilibrium, skilled workers will be indifferent between being unemployed and being in the informal sector, U = H, which implies that

$$F - U = F - H = \frac{w_h - b}{m - \lambda m}$$

Substituting for F-U then yields

$$w_f = \frac{(r + \delta + m(\theta))w_h - (r + \delta + \lambda m(\theta))b}{(1 - \lambda)m(\theta)} \tag{3.1}$$

The formal sector firms post vacancies and incur a cost, c, from searching for the appropriate match. When a worker is matched and production takes place, the firm receives the output,

$P_F$, less the wage given to the worker. The value of opening a formal sector vacancy ($V$) is then given by:

$$rV = -c + \frac{m(\theta)}{\theta}[J - V]$$

The value of an active job for the formal sector firm (J) is then

$$rJ = P_F - w_f + \delta[V - J]$$

In equilibrium, free entry ensures, V=0, in the formal sector implies

$$J = \frac{c\theta}{m(\theta)}$$

The value of the active job then becomes

$$J = \frac{P_F - w_f}{r + \delta}$$

From the above two equations, we have

$$w_f = P_F - \frac{(r + \delta)c\theta}{m(\theta)} \tag{3.2}$$

The match surplus must be allocated between workers and firms and we assume that the formal sector wage is determined by Nash bargaining with $\eta$ being the bargaining power of workers.

$$\max_w [F - U]^\eta \, [J - V]^{1-\eta}$$

We then get the formal sector wage to be equal to

$$w_f = \eta(P_F + c\theta) + (1 - \eta)b \tag{3.3}$$

There is a rural sector that has only unskilled workers producing a rural good by a Cobb-Douglas production function. Rural workers only consume rural informal sector output. The rural sector only uses unskilled workers to produce its output. It is perfectly competitive and characterized by full employment. The production function $R(x)$ is assumed to satisfy $R^{'} > 0$ and $R^{''} < 0$ where x is the number of unskilled workers employed in the rural sector. They earn their average productivity, $w_a$. The unskilled workers can work in the rural sector and earn $w_a$ plus the government assistance (z) or can work for the urban informal sector. z is the policy variable. In equilibrium, the unskilled workers are indifferent between working in rural areas and urban areas. So, we have

$$w_a + z = w_l \tag{3.4}$$

Markets for both skilled and unskilled workers should clear in equilibrium. $p$ is the fraction of workforce who are unskilled and 1-$p$ is the fraction who are skilled. Thus,

$$x + l = p$$

and

$$h + e + u = 1 - p$$

where u is the unemployment rate of skilled workers and e is formal sector employment. In steady state, flows into and out of formal sector employment should be equal.

$$m(\theta)u + \lambda m(\theta)h = \delta[1 - p - u - h] \tag{3.5}$$

Figure 3.6: Flow Diagram

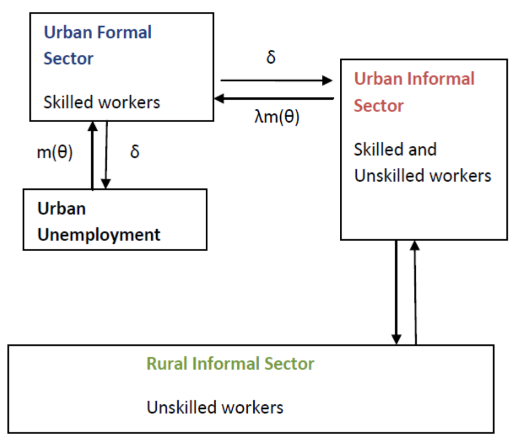



Figure 3.6 summarizes the labor market described above. The whole model can be expressed in three equations after substituting all variables in terms of l, h and $\theta$.

$$w_l - w_a - z \ = 0 \tag{C1}$$

Putting (3.1) and (3.2) together, we get

$$(1-\lambda)m(\theta)P_F - (1-\lambda)(r+\delta)c\theta - (r+\delta+m(\theta))w_h + (r+\delta+\lambda m(\theta))b = 0 \qquad (C2)$$

Putting (3.2) and (3.3) together, we get

$$(1-\eta)P_F - c\theta(\frac{r+\delta}{m(\theta)}+\eta)-(1-\eta)b = 0 \qquad (C3)$$

**Definition 1** *A steady state equilibrium is a set of* $\{\theta^*, u^*, l^*, h^*, x^*, w_a^*, w_{H,}^*\ w_L^*,\ w_f^*\}$ *which satisfies equations C1, C2, C3, market clearing conditions for the skilled and the unskilled workers, wages of the urban informal sector being their marginal productivity, wages in formal sector being Nash bargained and the wages in rural areas being the average productivity of the unskilled workers.*

**Proposition 1** *Assuming that the matching function* $m(\theta)$ *is continuous, concave and monotonically increasing, the urban informal sector and the urban formal sector outputs are complements in the final production function and the skilled and the unskilled workers in the urban informal sector are complements, the model has a steady state equilibrium.*

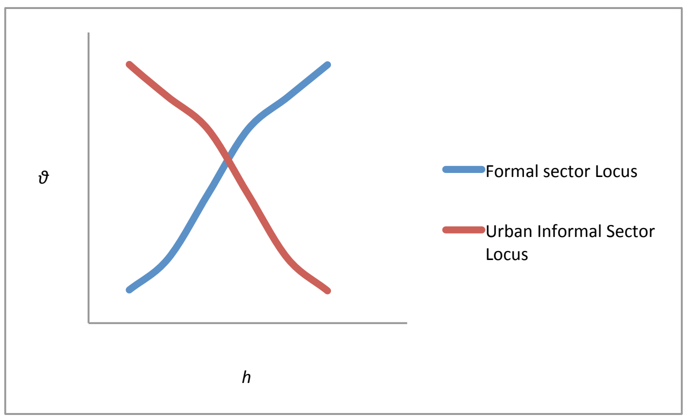
*Proof.* Using the market clearing condition for skilled workers and the steady state conditions, u, the unemployment rate, can be written as function of $\theta$ and h. C1, C2 and C3 can then be written as function of $\theta, l$ and h. Simplifying,

$$w_l(l,h) - w_a - z\ = 0 \qquad (C1)$$

$$(1-\lambda)m(\theta)P_F(l,h,\theta) - (1-\lambda)(r+\delta)c\theta - (r+\delta+m(\theta))w_h(l,h) + (r+\delta+\lambda m(\theta))b = 0 \qquad (C2)$$

$$(1-\eta)P_F(l,h,\theta) - c\theta(\frac{r+\delta}{m(\theta)}+\eta) - (1-\eta)b = 0 \qquad (C3)$$

From C1, since $h$ and $l$ are complements in the urban informal sector, $h$ is an increasing function of $l$. C2 and C3 are considered in the $(\theta, h)$ plane. C3 is the formal sector locus. A higher $h$ increases informal sector output which increases (decreases) the demand for formal sector output if the two outputs are complements (substitutes), so firms enter(leave) increasing (decreasing) the market tightness, $\theta$. C2 is the urban informal sector locus. An increase in $h$ will lower the wage of the skilled workers in the urban informal sector which will lower the value of being employed in the formal sector since the outside option is lower which will decrease $\theta$. It is also straightforward to see that as $\theta \longrightarrow 0, C3 \longrightarrow -(1-\eta)b$ and C2$\longrightarrow -\infty$ while as $\theta \longrightarrow \infty, C2 \longrightarrow -\infty$ and $C3 \longrightarrow -\infty$. Therefore, C3 is above C2 as $\theta \longrightarrow 0$. So there must be at least one intersection. Once $\theta$ and $h$ are determined, $l$ is calculated using C1. The rest of the variables are then calculated accordingly. The diagram below sketches the existence of equilibrium when $F$ and $I$ are complements.




### 3.6 Calibration

I calibrate the model to match the size of employment in the three sectors in 2004-2005 and then run a policy experiment and compare the model's prediction with the actual data from 2009-2010. Table 3.1 shows the target statistics while Table 3.2 shows the calibrated statistics.

Following World Bank (2010) and Bino (2008), I define the regular workers category in the survey as formal sector workers and the sum of self-employed and casual labor as the informal sector workers. People who have attained above secondary school are categorized as skilled while all others are categorized as unskilled for calibration purposes. The data comes from various tables of the NSSO survey on "Employment and Unemployment Situation in India," for 2004-2005 and 2009-2010 unless otherwise stated (Government of India, Ministry of Labor and Employment (2009) and Government of India, Ministry of Labor and Employment (2007)). The time frame is a year.

Following Acemoglu (2001), the production technology of the final good in urban areas is assumed to be Cobb-Douglas.

$$Y = AF^{\upsilon}I^{1-\upsilon}$$

Saha (2004) estimates formal sector output to be 40% and informal sector output to be 60% of the Indian GDP. Raveendran (2006) estimates that informal sector contributes to 57% of GDP. In this paper, I use the number from Saha (2004). Indiastat publishes CPI data for industrial workers (CPI-IW). It provides the data sectorwise. Since Bino (2008) reports that the formal sector is mainly concentrated in public administration, personal care and education, I calculate a formal sector price index by averaging the price indices of these sectors weighted by their share in total output. Similarly, Bino (2008) also reports that the informal sector is concentrated in manufacturing, construction and transport. Accordingly, we calculate an informal sector price index. The elasticity parameter, $\upsilon$ is calculated to be 0.39 ($\frac{P_f}{P_I} = \frac{\upsilon I}{(1-\upsilon)F}$).

The rural sector in the model includes both the self-employed and the casual labor who work in rural areas. I do not differentiate between people who work on their own land and people who work for others. I follow Gutierrez, Paci and Park (2009) in modeling the wage in the rural sector as the average product of labor where the production function is given by a Cobb-Douglas, $A_0x^{\gamma}$. I estimate the elasticity of rural labor ($\gamma$ ) by regressing log of real rural output on log of rural employment for the years 1990-2007. The data for rural output and employment come from Indiastat and World Development Indicator respectively. The regression result can be found in Table A.16. Table 49R of the 2004-2005

NSSO survey reports the wage of rural casual workers and then I estimate $A_0$ from the rural wage equation.

The urban informal sector is also assumed to be perfectly competitive which has both skilled and unskilled workers participating. The CES production function used for the baseline calibration is

$$I(l, h) = A_I[\alpha l^{\omega} + (1-\alpha)h^{\omega}]^{1/\omega}$$

In 2004, NSSO conducted a survey on manufacturing enterprises all over India (Schedule 62.2.2). It reports the number of owners, number of employees, education of the workers, wages, loans outstanding, expenses, receipts and sales for each enterprise. The survey also reports whether the entity is registered or not. Any company that is not registered is taken to be operating in the informal sector for this paper. I use the total value of goods and services sold as the total output of the company. Skilled and unskilled workers are defined as above. Using non-linear least square method, I estimate the above production function. $\alpha$ is estimated to be 0.89 and $\omega$ is estimated to be -0.48. The details of the estimation are provided in Table A.15. This provides evidence that skilled and unskilled workers are gross complements in the informal sector in India. This result is in contradiction to previous studies, namely Mello (2008) and Caselli and Coleman (2012). They find that unskilled and skilled workers are imperfect substitutes but they take overall employment and do not differentiate between the formal and informal sectors. In the Indian informal sector, firms are small, capital is scarce and the work that is being done by machine in developed countries is done by unskilled laborer. Skilled workers do work that does not require them to make their hands dirty. This can be a reason for unskilled workers and skilled workers to be complements in the Indian informal sector. This view has been advocated by Amaral and Quintin (2006). Table 24 and table 48 of the 2004-2005 NSSO reports on the "Employment and Unemployment Situation in India," report employment and wages of regular workers which I use as the formal sector wage and employment. The Ministry of Labor's annual report states that for the formal sector, the government requires that the unemployed get their wage for 120 days which is one fourth of a regular worker's wage in a year. Unemployment benefits, b, are therefore set to be $0.25w_f$. Table 76 of the NSSO report (2004) states that the flow from casual to regular job was 4 per 1000 person while that from unemployment was 33 per 1000. Assuming that this reflects amount of search done by individual, I set the search intensity parameter, $\lambda$, to be 0.125. The job separation rate, $\delta$, is set to be 0.166 as the survey finds that 166 persons out of 1000 left regular employment in 2004. Lastly, in line with the search and matching literature, a matching function of the form $m(\theta) = \theta^{1/2}$ is used with a bargaining parameter of 0.5.

Table 3.1: Target Statistics

| **Variable** | **Values in 2004** |
|---|---|
| Rural Sector Employment (x) | 0.59 |
| Urban Informal Sector Employment | 0.17 |
| Skilled Informal Sector Employment (h) | 0.03 |
| Unskilled Informal Sector Employment (l) | 0.14 |
| Formal Sector Employment (e) | 0.19 |
| Unemployment(u) | 0.05 |
| Rural sector productivity ($A_o$) | 34 |
| Elasticity parameter in rural production ($\gamma$) | 0.124 |
| Share parameter of skilled in urban informal production function (1-$\alpha$) | 0.11 |
| Elasticity parameter in urban informal production function ($\omega$) | -0.48 |
| Elasticity parameter in final production ($\upsilon$) | 0.39 |
| Interest rate (r) | 0.06 |
| Informal search Intensity Parameter ($\lambda$) | 0.125 |
| Job separation rate($\delta$) | 0.166 |
| Proportion of unskilled worker (p) | 0.73 |

Table 3.2: Calibrated Statistics

| **Variable** | **Value** |
|---|---|
| Informal technology parameter ($A_I$) | 86 |
| Formal technology parameter ($A_f$) | 0.08 |
| Cost (c) | 270 |
| Relative value of formal sector output ($P_F$) | 401 |
| Price of urban informal sector output ($P_I$) | 1 |
| Informal sector output (I) | 9.4 |
| Formal sector output (F) | 0.02 |
| Formal sector wage ($w_f$) | 96.5 |
| Informal skilled wage | 65.6 |
| Unskilled wage | 54 |

The other parameters are either collected directly from the NSSO reports or are assumed according to the literature. Calibration results suggest that the large informal sector in India might be a result of high entry costs, which is in line with previous findings.

## 3.7 Policy Implications

Since I find that skilled and unskilled workers are complements in the urban informal sector and formal and informal outputs are complements in India, in investigating policy effects, I assume $I_{hl} > 0$ and $Y_{FI} > 0$.

MNREGA can be financed by either foreign aid, borrowing or taxing. If the fund comes from foreign aid then no change happens to the model above. I also extend the model by allowing the government to collect the funds by taxing the formal sector firms.

**Corollary 1** *Suppose* $I_{hl} > 0$ *and* $Y_{FI} > 0$ *and* $(P_F - w_h) < \lambda(P_F - b)$, *then a wage subsidy financed by foreign aid to the rural sector leads to*

- *A decrease in unskilled employment in the urban informal sector and an increase in the wage, that is,* $\frac{dl}{dz} < 0$ *and* $\frac{dw_l}{dz} > 0$
- *An increase in skilled employment in the urban informal sector and a decrease in the wage, that is,* $\frac{dh}{dz} > 0$ *and* $\frac{dw_h}{dz} < 0$

- *A decrease in market tightness, output in the formal sector and the formal sector wage, that is,* $\frac{d\theta}{dz} < 0$, $\frac{dF}{dz} < 0$ *and* $\frac{dw_f}{dz} < 0$

- *An ambiguous effect on unemployment and informal output*

The proof of the corollary is given in the second appendix. Intuitively, a wage subsidy for the rural sector attracts unskilled workers from the urban informal sector towards the rural sector. As unskilled workers flow into the rural sector, wages of the unskilled workers in the urban informal sector increase. The decrease in unskilled labor in the informal sector decreases informal output. As output in the informal sector and the formal sector are complements, formal sector output falls which decreases employment of skilled labor in the formal sector. So, skilled workers take up informal sector jobs or become unemployed and look for other formal sector jobs. As skilled and unskilled workers are complements in the informal sector, as unskilled workers leave the informal sector, wages of skilled workers in that sector go down. As the relative value of the formal sector output falls, market tightness falls as new firms are discouraged to enter the market. There are two effects on unemployment- the fall in market tightness increases unemployment while the increase in skilled employment in the informal sector tends to decrease unemployment. The fall in market tightness and in the wage of skilled informal workers lower the effective bargaining power of skilled workers which leads to a lower formal sector wage.

**Corollary 2** *Suppose* $I_{hl} > 0$ *and* $Y_{FI} > 0$ *and* $(P_F - w_h) < \lambda(P_F - b)$*, then a wage subsidy to the rural sector, financed by a lump sum tax on the formal sector firms, leads to*

- *A decrease in unskilled employment in the urban informal sector and an increase in wage, that is,* $\frac{dl}{dz} < 0$ *and* $\frac{dw_l}{dz} > 0$

- *An increase in skilled employment in the urban informal sector and a decrease in wage, that is,* $\frac{dh}{dz} > 0$ *and* $\frac{dw_h}{dz} < 0$

- *A decrease in market tightness, output in the formal sector and the formal sector wage, that is,* $\frac{d\theta}{dz} < 0$, $\frac{dF}{dz} < 0$ *and* $\frac{dw_f}{dz} < 0$

- *An ambiguous effect on unemployment and informal output*

A lump-sum tax on the formal sector decreases market tightness more as new firms are discouraged from entering the formal sector. Therefore, skilled workers find jobs in the informal sector or become unemployed. The wage subsidy to the rural sector attracts unskilled labor from the urban informal sector to the rural sector but now there is a constraint on the amount of money that the government can transfer. Consequently, the rise in rural employment is lower under this scenario. The tax on the formal sector re-enforces the effects of wage subsidy on the formal sector firms by lowering employment even further.

### 3.8 Empirical Results

Using the calibrated model, I simulate the effects of the policy on the Indian economy and compare the results with the data from 2009-2010. The actual change reported is the simple difference-in-difference estimate between the treated and the non-treated districts. It should be mentioned here that the treatment was phased in according to some poverty criteria that my estimates do not capture. The administrative documents of MNREGA report that on average each rural household is getting only 40 days of work per year. The policy variable, z, is quantified by the average increase in annual real wage of a rural casual worker who works 40 days in the employment guarantee program and the rest of the days in non-public casual employment. MOSPI (Ministry of Statistics and Program Implementation) reports a price index of rural laborers and we use that price index to find the real wage. z is found to be 4.1. I then conduct two experiments - (1) Effects of the wage subsidy without taxing the formal sector firms, (2) Effects of the wage subsidy with a lump-sum tax on the formal sector firms. I also investigate the effects of the wage subsidy on wage inequality and welfare. Welfare is measured by a weighted average of lifetime income of workers in different sectors and inequality is measured by the wage ratio between the formal sector wage and the skilled urban informal sector wage, and by the wage ratio between the skilled and the unskilled urban informal sector wage.

### 3.8.1 Effect of wage subsidy without tax on formal sector

As predicted from the model, unskilled employment in the urban informal sector declined due to the wage subsidy as people move to the rural areas to get the benefits of the policy. While in reality, rural employment increased by 2.5%, the model predicts an increase of 2.3%. The value of formal sector output falls, which decreases market tightness. Skilled employment in the informal sector rises as people lose jobs in the formal sector and move towards the informal sector and unemployment. The model predicts formal sector employment to fall by 2% while in reality it fell by 10%. Urban informal sector employment was predicted to fall by 6.5% while the actual decrease was 8%. So the model predicts the movement of labor among sectors reasonably accurately, especially for the informal sector.

The model, however, lacks predictive power on wages although the prediction on the direction of the change is correct. The actual change in the wage of the rural informal sector net of the subsidy and the formal sector wage decreased a lot more than predicted by the model. This might be due to how the subsidy in being financed. If the scheme is financed by taxing the formal sector firms then the effect on the formal sector will be higher as seen by the following experiment.

Under this specification, the welfare of the unskilled workers rises while that of skilled workers fall. The policy can potentially reduce inequality in the informal sector but increase inequality between the urban formal and informal sectors. The increase in the unskilled wage and the decrease in the skilled wage reduce inequality in the urban informal sector. The decrease in the formal sector wage is less than the decrease in the skilled informal wage, increasing inequality between the formal and the informal sector for skilled workers. Thus, a development policy like MNREGA may reduce inequality and serve the targeted group as intended but it may have indirect adverse effect on the formal sector skilled workers which may reduce overall welfare.

Table 3.3: Effect of wage subsidy without tax on formal sector

| Variables | Predicted Change(%) | Actual change (%) |
|---|---|---|
| **Formal Sector** | | |
| Employment | -2.3 | -10 |
| Wage | -18 | -29 |
| Market Tightness | -20.84 | |
| **Urban Informal Sector** | | |
| Unskilled Employment | -9.7 | -9.2 |
| Skilled Employment | 8.6 | 1.6 |
| Total employment | -6.47 | -8 |
| Unskilled Wage | 5.5 | 10 |
| Skilled Wage | -19.8 | -12.6 |
| Average wage | 0.13 | 2.6 |
| **Rural Informal Sector** | | |
| Employment | 2.3 | 2.5 |
| Wage | -1.98 | -12 |
| Unemployment | 8.4 | 4 |

Table 3.4: Welfare

| | **Lifetime income** | |
|---|---|---|
| | **z=0** | **z=4.1** |
| All | 60.89 | 59.07 |
| Skilled | 79.61 | 64.11 |
| Unskilled | 53.96 | 56.99 |

Table 3.5: Inequality

| | **Inequality** | |
|---|---|---|
| | **z=0** | **z=4.1** |
| Between skilled and unskilled in urban informal sector | 1.21 | 1.02 |
| Between skilled in formal sector and urban informal sector | 1.48 | 1.51 |

### 3.8.2 Effect of wage subsidy on the rural sector when it is financed by taxing formal sector firms

Next I look into the effect of the wage subsidy on the rural sector when the subsidy is financed by a lump-sum tax on the formal sector firms. The effects on the formal sector are now greater. This might be due to the fact that the formal sector firms close down to avoid paying the taxes to finance MNREGA. The firms now prefer to enter the urban informal sector where they can evade paying taxes. The increase in skilled employment in the urban informal sector is higher than the model without tax. Overall, this model's predictions are closer to reality than the model with without tax. In reality, 0.5 percent of the Indian GDP is being spent on MNREGA in 2007. Under this scenario, the skilled informal wage falls less than the formal wage that leads to decrease in inequality within the urban informal sector and between the formal and the urban informal sector.

Table 3.6: Effect of wage subsidy on the rural sector when it is financed by taxing formal sector firms

| Variables | Predicted Change(%) | Actual change (%) |
|---|---|---|
| **Formal Sector** | | |
| Employment | -3 | -10 |
| Wage | -21 | -29 |
| Market Tightness | -20 | |
| **Urban Informal Sector** | | |
| Unskilled Employment | -10 | -9.2 |
| Skilled Employment | 9 | 1.6 |
| Total employment | -7 | -8 |
| Unskilled Wage | 6 | 10 |
| Skilled Wage | -19 | -12.6 |
| Average wage | 0.2 | 2.6 |
| **Rural Informal Sector** | | |
| Employment | 2.2 | 2.5 |
| Wage | -2 | -12 |
| Unemployment | 7 | 4 |

Table 3.7: Welfare

| | **Lifetime income** | |
|---|---|---|
| | **z=0** | **z=4.1** |
| All | 60.88 | 58.35 |
| Skilled | 79.61 | 61.98 |
| Unskilled | 53.96 | 57.05 |

Table 3.8: Inequality

| | **Inequality** | |
|---|---|---|
| | **z=0** | **z=4.1** |
| Between skilled and unskilled in urban informal sector | 1.21 | 1.01 |
| Between skilled in formal sector and urban informal sector | 1.48 | 1.47 |

### 3.8.3 Sensitivity Analysis

Due to the lack of data, I have assumed both the bargaining power and the matching function. In this section, I study how my results from the previous section change if the bargaining power and the matching technology are different from the baseline experiment.

#### Different bargaining power

In the previous section, I assume that the bargaining power is equal between the firm and the worker. If the bargaining power of the firm is higher ($\eta = 0.4$) than the formal sector workers, then the effect on the formal sector wage is higher than the effect in the baseline model. The firms pay the workers of the formal sector a low wage due to the wage subsidy given to the rural sector. As the firms can increase profit by lowering wage cost, less people lose their formal sector jobs. The effect on urban informal sector is lower and the effect on rural sector is higher than that in the baseline model.

Table 3.9: Effects of changing the bargaining power

| Variables | $\eta = 0.4$ | $\eta = 0.5$ | $\eta = 0.6$ |
|---|---|---|---|
| **Formal Sector** | | | |
| Employment | -1.4 | -2 | -2.8 |
| Wage | -18.4 | -18 | -17.8 |
| Market Tightness | -18.2 | -18 | -17.9 |
| **Urban Informal Sector** | | | |
| Unskilled Employment | -10 | -11 | -12 |
| Skilled Employment | 6 | 5 | 2 |
| Total employment | -7.8 | -8 | -9 |
| Unskilled Wage | 5.1 | 5.1 | 5.2 |
| Skilled Wage | -18 | -18 | -18.1 |
| Average wage | 0.2 | 0.28 | 0.3 |
| **Rural Informal Sector** | | | |
| Employment | 2.99 | 2.98 | 2.96 |
| Wage | -2.55 | -2.54 | -2.54 |
| Unemployment | 6 | 8.2 | 7 |

### Different matching technology

Increasing the frequency with which skilled workers find formal sector jobs decreases the effect on employment but increases the effect on wages. Unemployment increases more than before due to the higher increase in the value of unemployment than that in the value of working in the formal sector (Table 3.10). Overall, the result from the previous section holds, showing that the result is not sensitive to parameter specification of the matching function.

Table 3.10: Effects of changing the matching technology

| Variables | $m = 0.5\theta^{0.5}$ | $m = \theta^{0.5}$ | $m = 2\theta^{0.5}$ |
|---|---|---|---|
| **Formal Sector** | | | |
| Employment | -4 | -2 | -2 |
| Wage | -18 | -18 | -19 |
| Market Tightness | -20 | -18 | -17 |
| **Urban Informal Sector** | | | |
| Unskilled Employment | -13 | -11 | -10 |
| Skilled Employment | 2 | 5 | 7 |
| Total employment | -11 | -8 | -3 |
| Unskilled Wage | 5.1 | 5.1 | 5.3 |
| Skilled Wage | -17.8 | -18 | -19 |
| Average wage | 0.3 | 0.28 | 0.1 |
| **Rural Informal Sector** | | | |
| Employment | 3 | 2.98 | 2.5 |
| Wage | -2.56 | -2.54 | -2.2 |
| Unemployment | 7.2 | 8.2 | 8.6 |

### Different Elasticity Parameter in Urban Informal Sector Production

Although I have estimated the share and elasticity parameters in the production functions, I carry out some sensitivity checks in this section. If I vary the elasticity parameter of the urban informal production, the table below shows that as skilled and unskilled become more complement in the urban informal sector, the effects on employment in the urban informal sector and in the rural informal sector decrease (Table 3.11). However, the overall results do not change substantially.

Table 3.11: Effects of changing elasticity in the urban informal sector production

| Variables | $\omega = -0.46$ | $\omega = -0.48$ | $\omega = -0.50$ |
|---|---|---|---|
| **Formal Sector** | | | |
| Employment | -4.2 | -2 | -3.5 |
| Wage | -19 | -18 | -17 |
| Market Tightness | -22 | -20 | -20 |
| **Urban Informal Sector** | | | |
| Unskilled Employment | -10.6 | -9.7 | -9.6 |
| Skilled Employment | 13.6 | 8.67 | 8.46 |
| Total employment | -6.41 | -6.47 | -6.35 |
| Unskilled Wage | 5.4 | 5.5 | 5.65 |
| Skilled Wage | -21.5 | -19 | -19.6 |
| Average wage | -0.36 | 0.13 | 0.11 |
| **Rural Informal Sector** | | | |
| Employment | 2.5 | 2.98 | 2.3 |
| Wage | -2.2 | -1.97 | -1.94 |
| Unemployment | 8.2 | 8.2 | 8.2 |

### Different Share Parameter in Urban Informal Sector Production

Table 3.12 shows that decreasing the share parameter of the skilled workers in the urban informal sector production moves the model's predictions on all the sectors closer to the actual change but exaggerates the change in unemployment.

Table 3.12: Effects of changing share parameter in the urban informal section production

| Variables | $\alpha = 0.08$ | $a = 0.11$ | $\alpha = 0.14$ |
|---|---|---|---|
| **Formal Sector** | | | |
| Employment | -2.8 | -2 | -4.0 |
| Wage | -18.3 | -18 | -17.14 |
| Market Tightness | -20.86 | -20 | -20 |
| **Urban Informal Sector** | | | |
| Unskilled Employment | -11.2 | -9.7 | -8.5 |
| Skilled Employment | 6.36 | 8.67 | 9.6 |
| Total employment | -9.0 | -6.47 | -4.46 |
| Unskilled Wage | 4.8 | 5.5 | 6.03 |
| Skilled Wage | -19.9 | -19 | -18.9 |
| Average wage | 0.77 | 0.13 | -0.46 |
| **Rural Informal Sector** | | | |
| Employment | 3.02 | 2.98 | 1.87 |
| Wage | -2.57 | -1.97 | -1.6 |
| Unemployment | 9.4 | 8.2 | 6.8 |

### Other parameter sensitivity

Table 3.13 shows more sensitivity results. Decreasing labor elasticity in the rural sector, increasing formal sector output share in the production function of the final good, increasing job destruction rate and decreasing search intensity of skilled urban informal workers make the model's predictions closer to the actual changes.

Table 3.13: Effects of changing rural labor elasticity, share of formal sector in the final production, job destruction rate and search intensity

| | Base | $\gamma = 0.120$ | $\gamma = 0.128$ | $\upsilon = 0.3$ | $\upsilon = 0.5$ | $\delta = 0.160$ | $\delta = 0.172$ | $\lambda = 0.15$ | $\lambda = 0.25$ |
|---|---|---|---|---|---|---|---|---|---|
| e | -2 | -3.6 | -3.57 | -3.39 | -5.2 | -3.44 | -3.67 | -3.71 | -3.39 |
| $w_f$ | -18 | -18.04 | -18 | -18.43 | -18.09 | -18.07 | -17.98 | -18.11 | -17.94 |
| $\theta$ | -20 | -20.86 | -20.83 | -21.22 | -21.29 | -20.82 | -20.86 | -20.95 | -20.71 |
| l | -9.7 | -9.67 | -9.76 | -10.08 | -6.88 | -9.79 | -9.70 | -9.57 | -9.85 |
| h | 8.67 | 8.64 | 8.69 | 8.59 | 11.98 | 8.44 | 8.88 | 9.12 | 8.22 |
| l+h | -6.47 | -6.42 | -6.52 | -6.94 | -2.83 | -6.59 | -6.41 | -6.29 | -6.65 |
| $w_l$ | 5.57 | 5.55 | 5.63 | 5.83 | 5.05 | 5.57 | 5.64 | 5.68 | 5.51 |
| $w_h$ | -19.8 | -19.82 | -19.78 | -20.19 | -20.07 | -19.80 | -19.79 | -19.95 | -19.64 |
| x | 2.98 | 2.29 | 2.32 | 2.12 | 2.35 | 2.3 | 2.33 | 2.25 | 2.36 |
| $w_a$ | -1.97 | -1.97 | -1.98 | -1.82 | -2.02 | -1.97 | -1.99 | -1.93 | -2.02 |
| u | 8.2 | 8.38 | 8.42 | 8.8 | 5.77 | 8.47 | 8.14 | 8.26 | 8.35 |

One reason for the baseline predictions to differ from the actual change, especially of that in the urban sector may be due to the failure of the model to capture the global recession that hit India in late 2008.

## 3.9 Conclusion

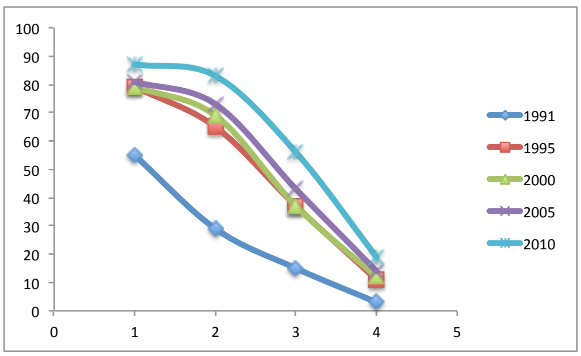
This paper builds a model of the Indian labor market and studies the implications of a recent development policy on labor markets, welfare and inequality. Using three sectors, I build a general equilibrium model where segmentation of the labor market arises due to matching frictions and heterogeneity of worker types. The model predicts that targeted job guarantee program like MNREGA may reallocate workers away from the urban informal sector but will lead to higher unemployment rate. It may increase the welfare of the unskilled rural people but may decrease the welfare of the skilled workers. Thus, policies like MNREGA might solve the problem for some portion of the labor force but at the cost of aggravating the problems for the others. Human capital development and decreasing institutional rigidities may help India to gain a more equitable distribution of the benefits of its economic growth.

The assumptions made in this paper are quite strong and relaxing them might be a direction for future work. The most obvious extension would be to study the effects of retrenchment and termination benefits policies on formal sector employment. The rural sector in India is not only more complex than presented here but also suffers from acute seasonal unemployment, which is not modeled here. The informal sector in India also constitutes of a large fraction of self-employed workers who act as unregulated micro-enterprises that is not modeled here. Previous research has already pointed out that modeling these extensions will need more sophisticated approach.

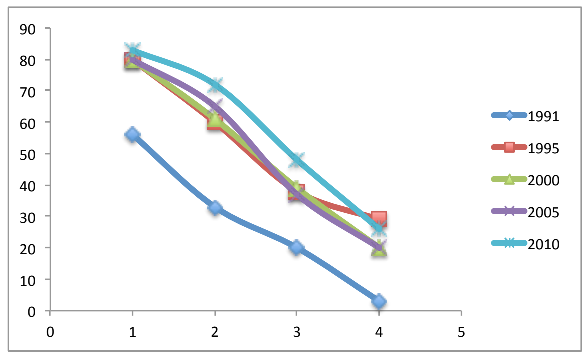
# Appendix A

## First Appendix

Figure A.1: Enrollment across education level and time




(a) Female




(b) Male

Note: x-axis gives the level of education: 1 - grade 1 to 5, 2 - grade 6 to 10, 3 - grade 11 to 12, 4 - grades above 12 and y-axis is the enrollment rate in percent

Table A.1: List of districts covered under FSSAP

| Division | Districts |
|---|---|
| Barishal | Barguna, Barishal, Bhola, Zhalokathi, Patuakhali, Pirozpur. |
| Chittagong | Brammanbaria,, Cox's Bazar, Rangamati, Khagrachari, Feni, Kumillah, Chadpur, Noakhali, Lakhipur, Chittagong. |
| Dhaka | Kishoreganj, Gazipur, Manikganj, Munsiganj, Moimansigh, Narayanganj, Narsingdi, Netrokona, Rajbari, Tangail, Gopalganj, Madaripur. |
| Khulna | Bagerhat, Chuadanga, Jessore, Zhinaidah, Khulna, Khustia, Satkhira, Magura, Narail. |
| Rajshahi | Jaipurhat, Noaga, Pabna, Rajshahi, Sirajganj, Bagura, Dinajpur, Gaibandha, Natore, Nilphamari, Panchgar, Rangpur, Thakurgaon. |
| Sylhet | Moulovibazar, Sunamganj, Sylhet, Habiganj. |

Table A.2: Cost of FSSAP according to academic year till 2002

| Year | Beneficiaries | Stipend | Tuition | Book allowance | SSC exams |
|---|---|---|---|---|---|
| 1994 | 902424 | 427806900 | 186678120 | 40216500 | 3095500 |
| 1995 | 1509314 | 710805540 | 306312300 | 55202750 | 44556500 |
| 1996 | 1767703 | 849110520 | 360791520 | 76953250 | 50283750 |
| 1997 | 2099202 | 1030094280 | 433308900 | 92437750 | 69307250 |
| 1998 | 2148266 | 1044951060 | 441051900 | 956228250 | 65444250 |
| 1999 | 2424813 | 1202064360 | 501567900 | 112929000 | 79163750 |
| 2000 | 2564577 | 1273237080 | 530621760 | 123706250 | 81892500 |
| 2001 | 2765061 | 1393575240 | 576205320 | 138201750 | 94429000 |
| 2002 | 2805293 | 1426109700 | 586907460 | 136325000 | 225668300 |

Table A.3: Sample size by sex and birth-cohorts

| Birth Year | Men | Female |
|---|---|---|
| 1950 | 968 | 545 |
| 1951 | 389 | 212 |
| 1952 | 3506 | 3737 |
| 1953 | 202 | 185 |
| 1954 | 1112 | 937 |
| 1955 | 656 | 4429 |
| 1956 | 883 | 331 |
| 1957 | 5026 | 541 |
| 1958 | 513 | 1641 |
| 1959 | 866 | 410 |
| 1960 | 2111 | 5395 |
| 1961 | 751 | 552 |
| 1962 | 5672 | 2042 |
| 1963 | 700 | 966 |
| 1964 | 2382 | 1370 |
| 1965 | 1595 | 6884 |
| 1966 | 1938 | 1370 |
| 1967 | 6773 | 6884 |
| 1968 | 994 | 864 |
| 1969 | 1366 | 1038 |
| 1970 | 3285 | 3044 |
| 1971 | 945 | 780 |
| 1972 | 5331 | 7556 |
| 1973 | 904 | 807 |
| 1974 | 3122 | 4218 |
| 1975 | 2369 | 2463 |
| 1976 | 2296 | 2650 |
| 1977 | 5273 | 7038 |
| 1978 | 427 | 385 |
| 1979 | 680 | 677 |
| 1980 | 479 | 526 |
| 1981 | 481 | 545 |
| 1982 | 1026 | 1114 |
| 1983 | 280 | 302 |
| 1984 | 455 | 471 |
| 1985 | 617 | 570 |

Table A.4: Impact Estimates using Difference-in-Difference-in-Difference(DDD) and Difference-in-Difference-in-Difference-in-Difference(DDDD) Framework

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | Education Attained | | | Labor Force Participation | | | Log Real Earnings | | |
| f | -2.50$^{**}$ | -2.56$^{**}$ | -2.56$^{**}$ | 0.07$^{**}$ | 0.08$^{**}$ | 0.08$^{**}$ | -0.81$^{**}$ | -1.03$^{**}$ | -1.03$^{**}$ |
| | (0.05) | (0.06) | (0.06) | (0.00) | (0.00) | (0.00) | (0.07) | (0.21) | (0.21) |
| T | 1.54$^{**}$ | 0.06 | 0.06 | 0.92$^{**}$ | 0.92$^{**}$ | 0.92$^{**}$ | 2.57$^{**}$ | 2.56$^{**}$ | 2.56$^{**}$ |
| | (0.08) | (0.08) | (0.08) | (0.00) | (0.00) | (0.00) | (0.06) | (0.12) | (0.12) |
| $S_i$ | -0.99$^{**}$ | -0.80$^{**}$ | | -0.00 | -0.01$^{**}$ | | -0.36$^{**}$ | -0.25$^{**}$ | |
| | (0.06) | (0.06) | | (0.00) | (0.00) | | (0.08) | (0.12) | |
| f*T | 0.38$^{**}$ | 0.71$^{**}$ | 0.72$^{**}$ | -0.90$^{**}$ | -0.93$^{**}$ | -0.93$^{**}$ | -0.04 | 0.15 | 0.15 |
| | (0.10) | (0.12) | (0.12) | (0.01) | (0.01) | (0.01) | (0.08) | (0.23) | (0.23) |
| $S_i$*T | 0.83$^{**}$ | 0.73$^{**}$ | | -0.03$^{**}$ | -0.03$^{**}$ | | 0.16$^{*}$ | 0.07 | |
| | (0.11) | (0.09) | | (0.01) | (0.01) | | (0.08) | (0.11) | |
| $S_i$*f | 0.85$^{**}$ | 0.93$^{**}$ | | 0.04$^{**}$ | 0.06$^{**}$ | | -0.19$^{**}$ | 0.17 | |
| | (0.06) | (0.08) | | (0.00) | (0.00) | | (0.09) | (0.26) | |
| $S_i$*T*f | 0.21$^{*}$ | 0.19 | | 0.02$^{*}$ | 0.03$^{**}$ | | 0.35$^{**}$ | -0.02 | |
| | (0.10) | (0.16) | | (0.01) | (0.01) | | (0.14) | (0.28) | |
| Age | -0.06$^{**}$ | -0.06$^{**}$ | -0.06$^{**}$ | -0.00$^{**}$ | -0.00$^{**}$ | -0.00$^{**}$ | -0.00 $^{**}$ | -0.00$^{**}$ | -0.00$^{**}$ |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Rural | -0.31$^{**}$ | 1.21$^{**}$ | 1.20$^{**}$ | 0.03$^{**}$ | 0.07$^{**}$ | 0.07$^{**}$ | -0.40$^{**}$ | -0.40$^{**}$ | -0.40$^{**}$ |
| | (0.04) | (0.03) | (0.03) | (0.00) | (0.00) | (0.00) | (0.03) | (0.02) | (0.02) |
| Household size | 0.02$^{**}$ | 0.02$^{**}$ | 0.02$^{**}$ | 0.00$^{**}$ | 0.00$^{**}$ | 0.00$^{**}$ | 0.00$^{**}$ | 0.00$^{**}$ | 0.00$^{**}$ |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Landownership | 0.03$^{**}$ | 0.03$^{**}$ | 0.03$^{**}$ | 0.00$^{**}$ | 0.00$^{**}$ | 0.00$^{**}$ | | | |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | | | |
| p | | -0.15$^{**}$ | -0.15$^{**}$ | | 0.01$^{**}$ | 0.01$^{**}$ | | -0.39$^{**}$ | -0.39$^{**}$ |
| | | (0.05) | (0.05) | | (0.00) | (0.00) | | (0.12) | (0.12) |
| p*f | | 0.09 | 0.09 | | -0.02$^{**}$ | -0.02$^{**}$ | | 0.37 | 0.37 |
| | | (0.07) | (0.07) | | (0.00) | (0.00) | | (0.24) | (0.24) |
| p*T | | 0.62$^{**}$ | 0.63$^{**}$ | | -0.02$^{**}$ | -0.02$^{**}$ | | 0.29$^{**}$ | 0.29$^{**}$ |
| | | (0.09) | (0.09) | | (0.01) | (0.01) | | (0.12) | (0.12) |
| p*T*f | | -0.39$^{**}$ | -0.39$^{**}$ | | 0.03$^{**}$ | 0.03$^{**}$ | | -0.26 | -0.26 |
| | | (0.15) | (0.15) | | (0.01) | (0.01) | | (0.26) | (0.26) |
| $S_i$*p | | -0.16$^{**}$ | | | -0.00 | | | -0.001 | |
| | | (0.07) | | | (0.00) | | | (0.08) | |
| $S_i$*p*T | | -0.12 | | | -0.02$^{**}$ | | | -0.38 $^{*}$ | |
| | | (0.10) | | | (0.01) | | | (0.20) | |
| $S_i$*p *T*f | | 0.31$^{*}$ | | | 0.03$^{**}$ | | | 0.48$^{**}$ | |
| | | (0.18) | | | (0.01) | | | (0.24) | |
| young | | | -1.06$^{**}$ | | | 0.03$^{**}$ | | | -0.46$^{**}$ |
| | | | (0.09) | | | (0.01) | | | (0.22) |
| middle | | | -0.67$^{**}$ | | | -0.02$^{**}$ | | | -0.20 |
| | | | (0.07) | | | (0.00) | | | (0.13) |
| young*f | | | 1.43$^{**}$ | | | 0.07$^{**}$ | | | 0.01 |
| | | | (0.11) | | | (0.01) | | | (0.38) |
| middle*f | | | 0.73$^{**}$ | | | 0.06$^{**}$ | | | 0.33 |
| | | | (0.08) | | | (0.01) | | | (0.29) |
| young*T | | | 0.71$^{**}$ | | | -0.11$^{**}$ | | | 0.08 |
| | | | (0.14) | | | (0.01) | | | (0.21) |
| middle*T | | | 0.73$^{**}$ | | | 0.00 | | | 0.09 |
| | | | (0.09) | | | (0.01) | | | (0.12) |
| young*p | | | 0.18$^{*}$ | | | 0.03$^{**}$ | | | 0.05 |
| | | | (0.10) | | | (0.01) | | | (0.13) |
| middle*p | | | -0.29$^{**}$ | | | 0.01$^{**}$ | | | -0.02 |
| | | | (0.07) | | | (0.00) | | | (0.08) |
| young*f*T | | | 0.37 | | | 0.01 | | | 0.114 |
| | | | (0.25) | | | (0.02) | | | (0.40) |
| middle*f*T | | | 0.12 | | | 0.03$^{**}$ | | | -0.16 |
| | | | (0.18) | | | (0.01) | | | (0.37) |
| young*f*p | | | -0.56$^{**}$ | | | 0.02$^{**}$ | | | -0.24 |
| | | | (0.14) | | | (0.01) | | | (0.37) |
| middle*f*p | | | 0.06 | | | 0.02$^{**}$ | | | -0.65$^{**}$ |
| | | | (0.11) | | | (0.01) | | | (0.29) |
| young*f*p*T | | | 0.52$^{*}$ | | | 0.07$^{**}$ | | | 0.42 |
| | | | (0.26) | | | (0.01) | | | (0.30) |
| middle*f*p*T | | | 0.35$^{*}$ | | | 0.02$^{*}$ | | | 0.57$^{**}$ |
| | | | (0.19) | | | (0.01) | | | (0.26) |
| Observations | 80738 | 139380 | 139380 | 80884 | 139713 | 139713 | 3636 | 7325 | 7325 |
| $R^2$ | 0.094 | 0.092 | 0.093 | 0.491 | 0.473 | 0.474 | 0.462 | 0.466 | 0.467 |

Standard errors in brackets
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$

Table A.5: Summary Statistics

| Destination | Age | Male |
|---|---|---|
| Bangladesh | 37.55 | 0.89 |
| | (13.58) | (0.32) |
| Saudi Arabia | 33.29 | 0.98 |
| | (8.95) | (0.13) |
| Kuwait | 34.20 | 0.97 |
| | (9.74) | (0.17) |
| UAE | 30.35 | 0.99 |
| | (8.30) | (0.11) |
| Bahrain | 32.88 | 0.96 |
| | (8.89) | (0.20) |
| Singapore | 28.04 | 1.00 |
| | (6.87) | (0.00) |
| Malaysia | 30.16 | 1.00 |
| | (7.87) | (0.05) |
| Gulf | 34.60 | 0.96 |
| | (9.82) | (0.18) |
| Asia | 31.63 | 0.99 |
| | (8.64) | (0.11) |
| West | 34.20 | 0.95 |
| | (9.17) | (0.22) |
| Africa | 30.00 | 0.98 |
| | (7.88) | (0.15) |
| Total | 37.52 | 0.89 |
| | (13.56) | (0.32) |

Mean is reported and standard error is given in parentheses.

Table A.6: Summary Statistics

| Destination | Years | Contractual | Search | Household Size | Cost |
|---|---|---|---|---|---|
| Saudi Arabia | 7.09 | 0.75 | 5.52 | 5.56 | 217.93 |
| | (5.22) | (0.44) | (8.63) | ( 2.90) | ( 391.33) |
| Kuwait | 8.42 | 0.81 | 6.47 | 5.70 | 191.41 |
| | (5.92) | (0.40) | (9.34) | (3.09) | (388.40) |
| UAE | 4.31 | 0.90 | 5.92 | 5.79 | 238.24 |
| | (5.34) | ( 0.29) | (8.64) | ( 3.08) | (559.17) |
| Bahrain | 6.22 | 0.68 | 6.00 | 5.83 | 259.62 |
| | (5.50) | (0.47) | (6.48) | (3.13) | (648.05) |
| Singapore | 3.99 | 0.94 | 5.32 | 5.20 | 316.80 |
| | (4.17) | (0.24) | (8.30) | (2.42) | (108.16) |
| Malaysia | 3.56 | 0.91 | 7.40 | 5.27 | 225.25 |
| | (4.23) | (0.28) | (7.42) | (2.65) | (253.49) |
| Gulf | 8.12 | 0.76 | 6.36 | 5.58 | 225.17 |
| | (7.60) | (0.43) | (10.88) | (3.09) | ( 582.85) |
| Asia | 6.44 | 0.64 | 5.93 | 5.57 | 318.31 |
| | (7.30) | (0.48) | (11.76) | (3.06) | (1056.77) |
| West | 7.95 | 0.28 | 7.48 | 5.79 | 604.00 |
| | (6.35) | (0.45) | (8.56) | (3.24) | (1175.99) |
| Africa | 3.86 | 0.67 | 9.52 | 5.77 | 272.28 |
| | (3.98) | (0.47) | (17.17) | (3.18) | (174.70) |

Mean is reported and standard error is given in parentheses. Cost of migration in thousands of Taka.

Table A.7: List of countries considered in aggregate

| Destination | Countries |
|---|---|
| Gulf | Qatar, Oman, Bahrain, Iran, Lebanon, Brunei, Iraq, Jordan, Libya |
| Asia | India, Nepal, Australia, New Zealand, Russia, Taiwan, China, Japan, Korea, Thailand, Indonesia |
| West | USA, UK, Canada Norway, Holland, Portugal, Poland, Sweden, Germany, Spain, Switzerland, Italy |
| Africa | South Africa, Egypt, Ivory Coast, Morocco, Nigeria |

Table A.8: Summary of the cell migration probability for stayers

| Education | Number of cells | Mean | Std. Dev. | 10th Percentile | 90th Percentile |
|---|---|---|---|---|---|
| Illiterate | 7542 | .8924932 | .0450269 | .8784173 | .9093108 |
| Primary | 7511 | .5805468 | .1503934 | .5140909 | .630609 |
| Secondary | 1389 | .6615507 | .1376838 | .6077586 | .8918919 |
| College | 812 | .7953116 | .0928788 | .7349398 | .8785047 |

Table A.9: Summary of the cell migration probability for movers

| **Education** | **Number of cells** | **Mean** | **Std. Dev.** | **10th Percentile** | **90th Percentile** |
|---|---|---|---|---|---|
| Illiterate | 933 | .0347416 | .0367263 | .0046064 | .0833333 |
| Primary | 7316 | .0568761 | .0260053 | .022827 | .0833333 |
| Secondary | 884 | .0557972 | .0297042 | .0159446 | .0833333 |
| College | 142 | .0476082 | .0356844 | .0082071 | .0833333 |

Table A.10: Corrected and uncorrected estimates of the return to college by country

| Destination | Uncorrected college return | Corrected college return | Hausman Test | Wald Test |
|---|---|---|---|---|
| Saudi Arabia | 0.52*** | 0.53*** | 0.02 | 27.4*** |
| | (4.66) | (3.42) | [0.88] | [0.00] |
| Kuwait | 0.83* | 0.96** | 1.73 | 8.37* |
| | (2.34) | (2.93) | [0.18] | [0.08] |
| UAE | 0.36** | 0.36** | 0.0003 | 10.5* |
| | (2.88) | (2.34) | [0.98] | [ 0.06] |
| Singapore | 0.41 | 0.59 | 0.34 | 8.17* |
| | (1.83) | (1.76) | [0.85] | [0.09] |
| Malaysia | 0.44 | 0.44 | 0.00 | 14.3*** |
| | (1.07) | (1.19) | [0.99] | [0.009] |
| Gulf | 0.64 | 0.63* | 0.00 | 7.15 |
| | (1.78) | (1.76) | [0.96] | [0.21] |
| Asia | 1.96*** | 1.88*** | 0.12 | 3.0 |
| | (6.15) | (5.23) | [0.72] | [0.70] |
| West | 0.30 | 0.5 | 1.73 | 5.51 |
| | (1.21) | (1.62) | [0.18] | [0.24] |
| Bangladesh | 1.063*** | 1.07*** | 0.01 | 8.88** |
| | (24.97) | (21.41) | [0.89] | [0.01] |

$t$ statistics in parentheses, p-value in square brackets

Table A.11: Corrected and uncorrected estimates of the return to secondary school by country

| Destination | Uncorrected secondary return | Corrected secondary return | Hausman Test |
|---|---|---|---|
| Saudi Arabia | 0.350*** | 0.342*** | 0.456 |
| | (8.02) | (12.06) | [0.499] |
| Kuwait | 0.487*** | 0.545*** | 1.99 |
| | (6.73) | (6.67) | [0.15] |
| UAE | 0.350*** | 0.347*** | 0.24 |
| | (7.89) | (6.93) | [0.62] |
| Singapore | 0.238* | 0.317** | 3.9* |
| | (2.38) | (2.86) | [0.05] |
| Malaysia | 0.246*** | 0.220*** | 1.66 |
| | (4.68) | (3.49) | [0.19] |
| Gulf | 0.288** | 0.317*** | 0.24 |
| | (3.26) | (3.68) | [0.62] |
| Asia | 1.1*** | 1.034*** | 0.47 |
| | (6.73) | (4.77) | [0.49] |
| West | 0.67*** | 0.67*** | 0.00 |
| | (19.65) | (20.19) | [0.99] |
| Bangladesh | 0.675*** | 0.672*** | 0.04 |
| | (18.93) | (18.42) | [0.84] |

$t$ statistics in parentheses, p-value in square brackets

Table A.12: Responsiveness of college relative to secondary school migration flows to differences in earnings and amenities

| Dependent variable: $\ln(p_k^h)$ - $\ln(p_k^f)$ | | | | | |
|---|---|---|---|---|---|
| | Uncorrected | Corrected | Corrected with amenities | | |
| Δ Earnings | 1.63*** | 0.94*** | 0.13*** | 0.34* | 0.21 * |
| | (13.89) | (8.27) | (8.4) | (2.14) | (2.1) |
| Distance | | | 0.76*** | 0.87*** | 0.19 |
| | | | (6.41) | (7.21) | (1.8) |
| Δ Unemployment | | | -0.18*** | -0.13*** | -0.05* |
| | | | (-6.14) | (-4.36) | (-2.1) |
| Δ Population density | | | -0.000059* | -0.000078** | -0.000088** |
| | | | (-2.46) | (-3.23) | (3.8) |
| Δ Temperature | | | -0.23*** | -0.28*** | -0.11*** |
| | | | (-14.76) | (-12.44) | (-5.15) |
| Cost | | | -0.000033*** | -0.000042*** | -0.000011*** |
| | | | (-9.94) | (-9.18) | (-3.14) |
| Δ Age Dependency Ratio | | | | -0.015** | -0.005*** |
| | | | | (-2.96) | (-7.05) |
| Migration Stock in 2000 | | | | | -0.03 *** |
| | | | | | (6.92) |
| [1em] Constant | 0.16*** | 0.088*** | 0.21*** | 0.22*** | 0.22*** |
| | (15.63) | (6.88) | (33.51) | (34.86) | (37.63) |
| $R^2$ | 0.16 | 0.06 | 0.26 | 0.27 | 0.29 |

$t$ statistics in parentheses. h is college education and f is secondary school.
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table A.13: Conditional Logit

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Corrected wage | 1.64*** | | | | | | | |
| | (14.98) | | | | | | | |
| | KSA | Kuwait | UAE | Singapore | Malaysia | Gulf | Asia | Europe |
| Secondary | 0.52*** | 0.65*** | 0.25** | 2.67*** | 0.44*** | 0.41** | -0.42* | 4.09*** |
| | (6.0) | (4.22) | (2.83) | (14.00) | (4.51) | (2.85) | (-1.77) | (13.52) |
| College | -0.57** | -0.72* | -0.80*** | 1.55*** | -1.77 | -1.65*** | -1.34*** | 4.17*** |
| | (-2.97) | (-1.95) | (-3.94) | (4.07) | (-0.8) | (-3.62) | (-3.48) | (13.22) |
| Log Likelihood: -26605.331 | | | | | | | | |
| Dissimilarity Coefficient: 1 | | | | | | | | |

$t$ statistics in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.14: Nested Logit with Amenities

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Corrected wage | 1.09*** | | | | | | | |
| | (15.02) | | | | | | | |
| Distance | -0.15*** | | | | | | | |
| | (-6.99) | | | | | | | |
| Unemployment | 0.23*** | | | | | | | |
| | (6.36) | | | | | | | |
| Population Density | 0.00 | | | | | | | |
| | (0.26) | | | | | | | |
| Temperature | -0.02*** | | | | | | | |
| | (-10.9) | | | | | | | |
| | KSA | Kuwait | UAE | Singapore | Malaysia | Gulf | Asia | Europe |
| Secondary | 1.3*** | 3.12*** | 0.03 | -5.76* | 18.78*** | 2.99*** | -19.9*** | 4.5*** |
| | (13.3) | (3.97) | (0.09) | (-2.67) | (8.39) | (3.84) | (-4.02) | (14.14) |
| College | 0.36 | 1.64 | -0.22 | -16.89 | -0.97 | 0.29 | 7.44 | 4.56*** |
| | (1.8) | (0.26) | (-2.25) | (-0.1) | (-0.14) | (1.43) | (1.63) | (12.93) |
| Log Likelihood: = -16857.161 | | | | | | | | |
| Dissimilarity Coefficient: 21.52 | | | | | | | | |
| LR test for IIA ($\tau = 1$): $\chi^2(1) = 521.1$ Prob> $\chi^2 = 0.0000$ | | | | | | | | |

$t$ statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.15: Estimation of CES function

| | Estimate | s.e | t-stat | p-value |
|---|---|---|---|---|
| ln $A_I$ | 6.9396 | 0.1507 | 46.05 | 0.00 |
| $\omega$ | -0.4782 | -0.1942 | 2.456 | 0.02 |
| $\alpha$ | 0.8854 | 0.2553 | 3.46 | 0.00 |
| R-squared | 0.1839 | Adjusted R-squared | 0.1688 | |
| Log-likelihood | -173.95 | Akaike criterion | 353.915 | |
| Schwarz criterion | 362.03 | | | |

Table A.16: Estimation of Rural Production Function

| | log(Rural Output) |
|---|---|
| log(rural employment) | 0.123*** |
| | (19.84) |
| cons | 1.690*** |
| | (201.09) |
| $N$ | 22 |

$t$ statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

# Appendix B

## Second Appendix

### B.0.1 Proof of Corollary 1

$\frac{dC1}{dl} = I_{ll} - w_{al} < 0$

$\frac{dC1}{dh} = I_{lh} > 0$

$\frac{dC1}{d\theta} = 0$

$\frac{dC2}{dl} = (1-\lambda)mP_{Fl} - (r+\delta+m)I_{hl} < 0$

$\frac{dC2}{dh} = (1-\lambda)mP_{Fh} - (r+\delta+m)I_{hh} > 0$

$\frac{dC2}{d\theta} = (1-\lambda)mP_{F\theta} + m'((1-\lambda)(P_F - c) - w_h + \lambda b) - (1-\lambda)(r+\delta)c < 0$

$\frac{dC3}{dl} = (1-\eta)P_{Fl} > 0$

$\frac{dC3}{dh} = (1-\eta)P_{Fh} > 0$

$\frac{dC3}{d\theta} = (1-\eta)P_{F\theta} - \eta c - (r+\delta)c\frac{d}{d\theta}(\theta/m)$

By Implicit Function Theorem, we get,

$$\begin{bmatrix} \frac{dC1}{dl} & \frac{dC1}{dh} & 0 \\ \frac{dC2}{dl} & \frac{dC2}{dh} & \frac{dC2}{d\theta} \\ \frac{dC3}{dl} & \frac{dC3}{dh} & \frac{dC3}{d\theta} \end{bmatrix} \begin{bmatrix} \frac{dl}{dz} \\ \frac{dh}{dz} \\ \frac{d\theta}{dz} \end{bmatrix} = \begin{bmatrix} -\frac{dC1}{dz} \\ -\frac{dC2}{dz} \\ -\frac{dC3}{dz} \end{bmatrix}$$

Determinant of the Jacobian $= |J| = \frac{dC1}{dl}(\frac{dC2}{dh}\frac{dC3}{d\theta} - \frac{dC2}{d\theta}\frac{dC3}{dh}) - \frac{dC1}{dh}(\frac{dC2}{dl}\frac{dC3}{d\theta} - \frac{dC2}{d\theta}\frac{dC3}{dl})$

$(\frac{dC2}{dl}\frac{dC3}{d\theta} - \frac{dC2}{d\theta}\frac{dC3}{dl}) > 0$

$(\frac{dC2}{dh}\frac{dC3}{d\theta} - \frac{dC2}{d\theta}\frac{dC3}{dh}) = (1-n)[(1-\lambda)(r+\delta)c - m'((1-\lambda)(P_F - c) - w_h + \lambda b)]P_{Fh} - (1-\eta)(r+\delta+m)I_{hl}P_{F\theta} + (r+\delta)c\frac{d}{d\theta}(\theta/m)((r+\delta+m)I_{hl} - (1-\lambda)mP_{Fh}) > 0$

$|J| < 0$

$\frac{dl}{dz} = \frac{(\frac{dC2}{dh}\frac{dC3}{d\theta} - \frac{dC2}{d\theta}\frac{dC3}{dh})}{|J|} < 0$ $\frac{dw_h}{dz} = I_{hh}\frac{dh}{dz} < 0$ $\frac{du}{dz} = \frac{du}{dh}\frac{dh}{dz} + \frac{du}{d\theta}\frac{d\theta}{dz}$ which is ambiguous.

$\frac{dh}{dz} = \frac{\frac{dC3}{dl}\frac{dC2}{d\theta} - \frac{dC2}{dl}\frac{dC3}{d\theta}}{|J|} > 0$ $\frac{dw_l}{dz} > 0$

$\frac{d\theta}{dz} = \frac{(\frac{dC2}{dh}\frac{dC3}{dl} - \frac{dC2}{dl}\frac{dC3}{dh})}{|J|} < 0$ $\frac{dF}{dz} = \frac{dF}{dh}\frac{dh}{dz} + \frac{dF}{d\theta}\frac{d\theta}{dz} < 0$

### B.0.2 Proof of Corollary 2

With tax on the formal sector firms, the value of an active job for the firms,

$$rJ = P_F - w_f - t + \delta(V - J)$$

where t is the tax.

Government balance budget: t(1-p-u-h) = z(p-l)

Substituting then gives us

$$t = \frac{z(p-l)(m(\theta)+\delta)}{(1-p-h(1-\lambda))m(\theta)}$$

t$'(l) = \frac{-z(m(\theta)+\delta)}{(1-p-h(1-\lambda))m(\theta)} < 0$

t$'(h) = \frac{z(p-l)(m(\theta)+\delta)(1-\lambda)m(\theta)}{((1-s-h(1-\lambda))m(\theta))^2} > 0$

t$'(\theta) = \frac{-\delta z(p-l)(1-s-h(1-\lambda))m'(\theta))}{((1-s-h(1-\lambda))m(\theta))^2} < 0$

The system of equations become,

*Proof.*

$$w_L(l,h) \text{ - } w_a - z = 0 \tag{B.1}$$

$$(1-\lambda)m(\theta)(P_F(l,h,\theta)-t(l,h,\theta)-c)-(1-\lambda)(r+\delta)c\theta-(r+\delta+m(\theta))w_h(l,h)+(r+\delta+\lambda m(\theta))b=0 \tag{B.2}$$

$$(1-\eta)(P_F(l,h,\theta)-t(l,h,\theta)-c)-c\theta(\frac{r+\delta}{m(\theta)}+\eta)-(1-\eta)b=0 \tag{B.3}$$

$\frac{dD2}{dl}=(1-\lambda)m(P_{Fl}-t_l)-(r+\delta+m)I_{hl}<0$

$\frac{dD2}{dh}=(1-\lambda)m(P_{Fh}-t_h)-(r+\delta+m)I_{hh}>0$

$\frac{dD2}{d\theta}=(1-\lambda)m(P_{F\theta}-t_\theta)+m'((1-\lambda)(P_F-c)-w_h+\lambda b)-(1-\lambda)(r+\delta)c<0$

$\frac{dD3}{dl}=(1-\eta)(P_{Fl}-t_l)>0$

$\frac{dD3}{dh}=(1-\eta)(P_{Fh}-t_h)>0$

$\frac{dD3}{d\theta}=(1-\eta)(P_{F\theta}-t_\theta)-\eta c-(r+\delta)c\frac{d}{d\theta}(\theta/m)$

By Implicit Function Theorem, we get,

$$\begin{bmatrix} \frac{dD1}{dl} & \frac{dD1}{dh} & 0 \\ \frac{dD2}{dl} & \frac{dD2}{dh} & \frac{dD2}{d\theta} \\ \frac{dD3}{dl} & \frac{dD3}{dh} & \frac{dD3}{d\theta} \end{bmatrix} \begin{bmatrix} \frac{dl}{dz} \\ \frac{dh}{dz} \\ \frac{d\theta}{dz} \end{bmatrix} = \begin{bmatrix} -\frac{dD1}{dz} \\ -\frac{dD2}{dz} \\ -\frac{dD3}{dz} \end{bmatrix}$$

Determinant of the Jacobian $=|J|=\frac{dD1}{dl}(\frac{dD2}{dh}\frac{dD3}{d\theta}-\frac{dD2}{d\theta}\frac{dD3}{dh})-\frac{dD1}{dh}(\frac{dD2}{dl}\frac{dD3}{d\theta}-\frac{dD2}{d\theta}\frac{dD3}{dl})$

$(\frac{dD2}{dl}\frac{dD3}{d\theta}-\frac{dD2}{d\theta}\frac{dD3}{dl})>0$

$(\frac{dD2}{dh}\frac{dD3}{d\theta}-\frac{dD2}{d\theta}\frac{dD3}{dh})=(1-n)[(1-\lambda)(r+\delta)c-m'((1-\lambda)(P_F-t-c)-w_h+\lambda b)](P_{Fh}-t_h)-(1-\eta)(r+\delta+m)I_{hl}(P_{F\theta}-t_\theta)+(r+\delta)c\frac{d}{d\theta}(\theta/m)((r+\delta+m)I_{hl}-(1-\lambda)m(P_{Fh}-t_h))>0$

$|J|<0$

$\frac{dl}{dz}=\frac{(\frac{dD2}{dh}\frac{dD3}{d\theta}-\frac{dD2}{d\theta}\frac{dD3}{dh})}{|J|}<0$ $\quad \frac{dw_h}{dz}=I_{hh}\frac{dh}{dz}<0$ $\quad \frac{du}{dz}=\frac{du}{dh}\frac{dh}{dz}+\frac{du}{d\theta}\frac{d\theta}{dz}$ which is ambiguous.

$$\frac{dh}{dz} = \frac{\frac{dD3}{dl}\frac{dD2}{d\theta} - \frac{dD2}{dl}\frac{dD3}{d\theta}}{|J|} > 0 \qquad \frac{dw_l}{dz} > 0$$

$$\frac{d\theta}{dz} = \frac{(\frac{dD2}{dh}\frac{dD3}{dl} - \frac{dD2}{dl}\frac{dD3}{dh})}{|J|} < 0 \qquad \frac{dF}{dz} = \frac{dF}{dh}\frac{dh}{dz} + \frac{dF}{d\theta}\frac{d\theta}{dz} < 0$$

# Bibliography

D. Acemoglu. Good Jobs and Bad Jobs. *Journal of Labor Economics*, 19(1), 2001.

A. Alam, J. E. Baez, and X. V. Del Carpio. Does Cash for School Influence Young Women's Behavior in the Long Term: Evidence from Pakistan. Unpublished manuscript, 2011.

J. Albrecht, S. Vroman, and L. Navarro. The Effects of Labor Market Policies in an Economy with an Informal Sector. *Economic Journal*, 2009.

P. Amaral and E. Quintin. A Competitive Model of the Informal Sector. *Journal of Monetary Economics*, 53(7), 2006.

M. Azam. The Impact of Indian Job Guarantee Scheme on Labor Market Outcomes: Evidence from Natural Experiment. *IZA Discussion Paper*, (6548), 2012.

Banbeis. http://www.banbeis.gov.bd, 2012.

J. R. Behrman, S. W. Parker, and P. E. Todd. Long term Impacts of Opprtunidades Conditional Cash Transfer Program on Rural Youth in Mexico. Unpublished manuscript, 2005.

M. V. Belot and T. J. Hatton. Immigration Selection in the OECD. *Scandinavian Journal of Economics*, 2013.

S. Bertoli, J. F. Moraga, and F. Ortega. Crossing the Border: Self-selection, Earnings and Individual Migration Decisions. *Journal of Development Economics*, 2012.

P. Bino. Indian Labour Market Report, 2008.

P. Carneiro and J. J. Heckman. The Evidence of Credit Constraints in Post Secondary Education. *Economic Journal*, 112(482), 2002.

F. Caselli and A. Coleman. The World Technology Frontier. *American Economic Review*, 96, 2012.

S. Chakrabarti and A. Kundu. Formal-Informal SectorsÕ Conflict: A Structuralist Framework for India. *Journal of Economic Development*, 34(2), 2009.

M. Clemens, C. E. Montenegro, and J. G. Pritchett. The Place Premium: Wage Differences for Identical Workers across the US Border. *Center for Global Development Working Paper*, 2008.

G. B. Dahl. Mobility and the Return to Education: Testing a Roy Model with Multiple Markets. *Econometrica*, 70(6), 2002.

E. Duflo. Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *The American Economic Review*, 91(4), 2001.

J. Farrington and D. Sjoblom. Employment Guarantee Act: Will it Reduce Poverty and Boost the Economy? *Project Briefing*, 2008.

G. Fields. Rural-Urban Migration, Urban Unemployment and Underemployment, And Job Search Activity in LDCs. *Journal of Development Economics*, 2, 1975.

N. Fuwa. The Net Impact of the Female Secondary School Stipend Program in Bangladesh. Unpublished manuscript, 2001.

Government of India, Ministry of Labor and Employment. Employment and Unemployment Situation in India 2004-2005, 2007.

Government of India, Ministry of Labor and Employment. Employment and Unemployment Situation in India 2009-2010, 2009.

J. Grogger and G. H. Hanson. Income Maximization and the Selection and Sorting of International Migrants. *Journal of Development Economics*, 95(1), 2011.

C. Gutierrez, P. Paci, and B. Park. Quantifying the Impact of the Financial Crisis on Employment and Wages in a Multi-Sector Labor Market. 2009. World Bank.

C. Harmon and I. Walker. Estimates of the Economic Returns to Schooling for the United Kingdom. *The American Economic Review*, 85(1), 1995.

J. R. Harris and M. P. Todaro. Migration, Unemployment and Development: A Two-Sector Analysis. *American Economic Review*, 60, 1970.

R. Heath and A. M. Mobarak. Does Demand or Supply Constrain Investments in Education: Evidence from Garment Sector Jobs in Bangladesh. Unpublished manuscript, 2012.

J. J. Heckman. Sample Selection Bias as a Specification Error. *Econometrica*, 47, 1979.

G. W. Imbens and T. Lemieux. Regression Discontinuity Designs: A guide to Practice. *Journal of Econometrics*, 2007.

C. Imbert and J. Papp. Equilibrium Distributional Impacts of Government Employment Programs: Evidence of India's Employment Guarantee, 2012. URL `http://www.parisschoolofeconomics.eu/docs/imbert-clement/2012-03-19-pse-working-paper-equilibrium-distributional-impacts-of-government-programs-imbert-papp.pdf`.

S. Khandker, M. Pitt, and N. Fuwa. Subsidy to Promote Girl's Secondary Education: The Female Stipend Program in Bangladesh. *MRPA*, (23688), 2003.

R. A. Mahmood. Overseas Migration for Employment: The Bangladesh Experience. *Journal of Economic Cooperation Among Islamic Countries*, 1993.

W. Mahmud. The Impact of Overseas Labour Migration on the Bangladesh Economy: A Macro-economic Perspective. *To the Gulf and Back: Studies on the Economic Impact* *of Asian Labour Migration*, 1989.

J. A. Maluccio and R. Flores. Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan Red de Proteccion Social. Unpublished manuscript, 2005.

D. Mazumder. Segmented Labor Markets in LDCs. *The American Economic Review*, 1983.

D. L. McFadden. Econometric Analysis of Qualitative Response Models. *Handbook of Econometrics*, II, 1984.

D. J. McKenzie and H. Rapport. Self-Selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks. *The Review of Economics and Statistics*, 92(4), 2010.

M. Mello. Skilled Labor, Unskilled Labor, and Economic Growth. *Economic Letters*, 100, 2008.

P. Oreopoulos. The Compelling Effects of Compulsory Schooling: Evidence from Canada. *Canadian Journal of Economics*, 39(1), 2006.

F. Ortega and G. Peri. The Causes and Effects of International Labor Mobility: Evidence from OECD countries 1980-2005. *NBER Working Paper Series No. 14833*, 2009.

M. M. Rahman. Does Labour Migration bring about Economic Advantage? A Case of Bangladeshi Migrants in Saudi Arabia. *ISAS Working Paper*, 2011a.

M. M. Rahman. Recruitment of Labor Migrants for the Gulf States: The Bangladeshi Case. *ISAS Working Paper*, 2011b.

G. Raveendran. Estimation of Contribution of Informal Sector to GDP, 2006.

A. D. Roy. Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, 3(2), 1951.

V. Saha. Contribution of Informal Sector and Informal Employment in Indian Economy, 2004.

M. Satchi and J. Temple. Labor Markets and Productivity in Developing Countries. *Review of Economic Dynamics*, 12, 2009.

A. T. Schurmann. Review of Bangladesh Female Secondary Stipend Project Using Social Exclusion Framework. *Journal of Health, Population and Nutrition*, 27(4), 2009.

A. K. Sen. Peasants and Dualism with or without Surplus Labor. *The Journal of Political Economy*, 74(5), 1966.

A. N. Sharma. Flexibility, Employment and Labor Market Reforms in India. *Economic and Political Weekly*, 2006.

V. Soares, F. R. P. Ribas, and G. I. Hirata. Achievements and Shortfalls of a Conditional Cash Transfers: Impact Evaluation of Paraguay's Tekopora Programme. Unpublished manuscript, 2008.

World Bank. *India's Employment Challenge: Creating Jobs, Helping Workers*. Oxford, 2010.

World Bank. *Migration and Remittances Factbook*. 2011.

World Bank. Poverty Assessment Report for Bangladesh: 2010, 2012.

World Economic Forum. The Global Gender Gap Report: 2012, 2012.

Y. Zenou. Job Search and Mobility in Developing Countries: Theory and Policy Implications. *Journal of Development Economics*, 86, 2008.

L. Zimmerman. Labor Markets Impacts of a Large-Scale Public Works Program: Evidence from the Indian Employment Guarantee Scheme. *IZA Discussion Paper*, (6858), 2012.