# IMPACT OF MIGRATION ON JOB SATISFACTION, PROFESSIONAL EDUCATION AND THE INFORMAL SECTOR

By

ILHOM ABDULLOEV

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Economics

Written under the direction of

Professor Ira Gang

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

OCTOBER, 2013

**ABSTRACT OF THE DISSERTATION**

**Impact of Migration on Job Satisfaction, Professional Education and the Informal Sector**

**By ILHOM ABDULLOEV**

**Dissertation Director:**

**Professor Ira Gang**

This dissertation focuses on three aspects of the economic life of migrants' relatives who remained in the source country. In the second chapter, we argue that migration increases the job dissatisfaction of migrants' relatives who work in the source country. The family's migration experience allows its members to construct expectations on earnings from migration using information either from the size of remittances or directly from migrants. If their expected earnings from migration greatly exceed their current wages in the source country, migrants' relatives become more dissatisfied with their current jobs. In the third chapter, we argue that migration has both positive and negative effects on education. In estimating the positive impact of remittances we suggest controlling for migration's side effects by including a dummy variable indicating whether there is a migrant in the household. This way, holding remittances constant, the coefficient on the dummy variable captures other negative side effects of migration as a change in the intercept between households with migrants and households without migrants. In the fourth chapter, we discuss how migration reduces informality in the source country: migrants' incomes in new locations and income earned in the home informal economy become an imperfect trade-off. It is also because professional workers have more opportunities to engage in informal activities enabling them to forgo migration, but low-skilled non-professionals do not. We offer concluding comments in the final chapter.

**Acknowledgement**

I would like to express my sincere appreciation and deepest gratitude to my academic advisor, Professor Ira Gang, for his understanding, concern and encouragement during the entire time of my study at Rutgers University. His support, counseling and guidance inspired me throughout my researches, and were important in promoting me as an academic researcher. I believe that without his persistent help and precious advices this dissertation would not be possible.

I would like to thank sincerely Professor John Landon-Lane for guiding me through advanced theories of statistics and the Bayesian econometrics. His practical advices and academic support were extremely helpful in developing my research ideas.

I also owe my thankfulness to Professor Hiroki Tsurumi for introducing me into the Bayesian econometrics and Professor Roger Klein for guiding me thoroughly through the semiparametric econometrics. I also want to thank Professor Alexander Danzer, my outside member, for graciously giving up his precious time to add his insights to my work.

I thank my dearest parents for their love, care and encouragement, my beloved wife for her understanding and patience, and my son, Jovid, for cheering me up.

I am grateful to the Open Society Institute for providing me funding to accomplish my doctoral study at Rutgers University.

*Research of this dissertation was supported in part by the Doctoral Fellowship Program, which is funded and administered by the Open Society Institute (OSI). The opinions expressed herein are the author's own and do not necessarily express the views of OSI.*

## Dedication

To my parents, Professor Abdumanon Abdulloev and Doctor Olga Nuriddinova.

## Table of Contents

## List of Tables

## List of Illustrations

## Part I. Introduction

Having a family member who currently lives and works abroad affects the social and economic behavior of families and their members in the migrant sending (source) countries. The absence of a migrant along with remittances sent to his family change attitudes of his family members who have remained in the source country towards their employment and schooling. Such aspects of international migration can be either welfare promoting or worsening.

Three different models on the effect of migration on such behavior changes of migrants' relatives in employment and education in the migrant's source country are considered in this dissertation. In the second chapter, we study how migration affects on the job satisfaction of migrant's family members remained in the source country. A migrant's relative who remains and works in the source country accesses information on the wage distribution in the destination country through either remittances that the family receives or via the migrant directly. Once there information is available, the relative builds her expectation on possible earnings as if she joined her migrant relative. Furthermore, since the family is involved in migration, the members know how to reduce migration related costs, which allows them to very closely predict the earnings they ultimately receive. As a result, the larger the difference between that what they receive now and their expected earnings from migration, the more dissatisfied current working migrant relatives will be.

In the third chapter, we study how migration impacts decisions of migrants' relatives on acquiring the professional education in the source country. We suggest separating the effect of remittances from the other side effects of migration on decisions about acquiring education. The main issue in previous studies was that they did not separate the effect of remittances from other migration side effects. These studies' major finding is that remittances help to increase the

likelihood of staying in schools of migrant family members since they lift budget constraints, so migrant families can “afford” education, and hire others to do housework. However, there are other side effects of migration too which in general reduce schooling. For example, the migrant’s absence causes less parental input in children's education, while restructuring household labor supply may require students to work, or children expecting to migrate with their migrant parents, could drop-out of school. If these side effects of migration matter, then ignoring them in the estimation of the effects of remittances biases its estimate, because of their high correlation. Our strategy is to separate the effect of remittances from other side effects of migration on the decisions about education by including a dummy variable on whether a household has a current migrant along with remittances. Such a variable would explain the change in intercept between the households with migrants and without due to the side effects of migration keeping remittances constant.

In the fourth chapter, we are interested in studying how migration is related to informal sector activities, whether it complements the informal sector, or substitutes. Migration is related to the informal activities in different directions depending on whether one studies this relationship from the migrant destination or source perspective. The current economic literature studies the effect of migration on the informal sector from the perspective of the destination country. From this perspective, the relationship between migration and the informal sector is complementary: new coming migrants may have difficulty in finding employment in formal work in the destination country, so many of them end up informally employed. However, the current literature does not discuss this relationship from the perspective of the migrant’s source country.

From the perspective of the migrant’s source country, the relationship between the informal sector and migration might be different. Firstly, in transitional and developing countries people with the high professional education have more opportunities for being involved in formal

sector income, while low skilled labor does not. The cost of migration in this case is higher for professional workers. There might be an income effect caused by remittances on the informal sector employment decisions. With higher income because of the receipt of remittances, migrant relatives who were working in informal sector might choose to consume more leisure by quitting informal work. Furthermore, migrant's remittances might help to improve the migrant family's financial conditions and, consequently, encourage family members to all together avoid the informal employment. In such cases, migration becomes a substitute for informal activities in the migrant's source country.

Due to different econometric issues, we applied three different estimation methods throughout the dissertation. In studying the relationship between the migration and informal sector, we used the Ordinary Least Squares. In studying the effect of migration on the individual job satisfaction we used both parametric and semiparametric estimation methods, while in the latter we were able to control for the endogeneity of the variable o interest. The final model of the effect of migration side effects and remittances on the decisions on professional education due to multiple endogeneity issue, we applied the Limited Information Bayesian Estimation based on Monte Carlo Markov Chain algorithms.

We have chosen as a country case Tajikistan, a transitional country which currently experiences a high level of international labor migration. Both remittances and migration play an important role for Tajikistan. The World Bank reports that the size of remittances sent by Tajik migrants reached one-third of the country's GDP in 2009 ranking Tajikistan as the world's most highly dependent country on remittances (World Bank, 2011).

After receiving its formal political independence from Russia, Tajikistan as other former Soviet republics underwent severe political and economic transformation. The civil war which

followed after the collapse of the Soviet Union caused severe impoverishment of its population. According to the World Bank, 53.5% of Tajikistan's population was living below its national poverty line even in 2007, after a decade from its national reconciliation. Different economic factors such as a lower natural resource endowment, its underdeveloped economy inherited from the Soviet government with a large share of its population employed in the agricultural sector, mainly in cotton fields, its landlockedness and many others, slowed down its economic growth until the rise of its international migration.

After the collapse of the Soviet Union, some countries because of their rich resource endowments along with effective implementation of economic reforms were able to achieve much higher rates of income growth than Tajikistan. In 2011, the GDP per capita for the Russian Federation the major country destination of Tajikistan's migrants was US$13,089 − fourteen times larger than Tajikistan's GDP per capita rate of US$935 (World Bank, 2013). Furthermore, the average real wages in Russian Federation were about 8.5 times higher than those in Tajikistan in August 2010. Under the new democratic government, international migration is no longer strictly controlled by the government, as it used to be during the Soviet rule. People can freely move not only within their own countries but also to some former Union's countries. Such large differential incomes between these two countries along with free entry travel policies for Tajikistan's citizens to Russia drive more citizens of the poorest former Soviet country − Tajikistan − into migration and made this country as the most remittance dependent country in the world.

For all three models, we used the data collected in 2007 for the World Bank Living Standard Survey (2007 WB LSS). This survey includes extensive questions on migration, education, health, labor market, housing, transfers and social assistance, subjective poverty and food security, as well as data for household's expenditure and income.

## Part II. Chapter 1. Impact of Migration on Job Dissatisfaction

### *1.1. Introduction*

Our focus in studying the employment-migration relationship is on the source country. We look at how the migration of one person to another (destination) country impacts the labor supply decisions of his relatives who remain in a source country. While all previous studies looked at the effect of remittances on the labor supply decisions of migrant relatives, in this paper we examine how the outmigration affects the job satisfaction of non-migrating relatives.

Studying the effect of migration on the labor supply is not new. The pioneering work on the effect of migration on labor supply of non-migrant family members is by Rodriguez and Tiongson (2001). They find a negative effect of remittances on the labor supply of the migrants' family members in urban Philippines: an additional US$40 of remittances per migrant family member decreases male and female labor participation by 0.3 and 0.2 percents respectively. Subsequent papers find further evidence of this negative effect of remittances on the labor supply. For example, Acosta (2006) finds that remittances received from international migrants reduce the likelihood of labor supply by children and women in migrants' families in El Salvador. Kim (2007) also finds the negative effect of remittances on the labor supply at both the individual and geographical cluster levels in Jamaica. Nguen and Purnamasari (2011) study the Indonesian data and find that migrant family members work 26 hours less per week than members of households without migrants; if migrant is male, his family members work 33 hours less than members of non-migrant households. Amuedo-Dorantes and Pozo (2006) reported that a 100 Peso remittance increase would reduce male formal sector employment by 32 hours per month in both urban and rural areas of Mexico, male self-employment by 11 hours per month in urban areas, female nonpaid employment by 6 hours per month, and female informal sector employment by 12 hours

per month. Cabegin (2006) studies migration from the Philippines, finding in families with wife-migrants that an annual increase in wives' earnings by 10,000 Pesos decreases the likelihood of having the full-time paid employment of their husbands by 12 percent more than men in non-migrant families. The same increase also leads to a rise in the likelihood of husbands being unemployed by 6 percent. In families with husband-migrants, the same size increase in husbands' earnings reduces the likelihood of full employment by their wives by 4 percent relatively to those in non-migrant households.

Why do remittances negatively affect individual labor supply decisions? Since remittances received from the migrant might have the same effect as that of the non-wage income in the individual (or family) utility maximization problem, there are two possible outcomes. Firstly, remittances could result in an interior solution in the labor supply problem, where the marginal rate of substitution between the consumption and leisure is equal to the real wage rate. Under this condition, if leisure is a normal good, then the increase in non-wage income reduces hours of work of migrant family members. Secondly, remittances might result in a corner solution to the labor supply problem when the marginal rate of substitution of consumption and leisure is greater than the wage rate. Since non-wage income raises individual budget constraints, it also increases individual reservation wages. Once individual reservation wages are increased to such level that they are higher than market wages, migrant family members would choose not to work (for detailed discussion of the effect of the non-wage income on the labor supply see Killingsworth (1983)).

However, as Rodriguez and Tiongson (2001) stated it is not entirely clear "whether migrants' remittances have a similar effect on labor supply as other nonlabor income" (p. 721). Due to the complexity of migration process there are different attributes that along with remittances influence labor supply decisions of migrant family members. Several authors discuss

these indirect effects of remittances and migration. Acosta (2006) mentions that the absence of the migrant along with the inflow of remittances might create positive externalities for neighbors of migrant families by relaxing the financial constraints they face as the migrant's family hires neighbors to do some work in their household to compensate migrant's absence. Kim (2007) hypothesizes that remittances are hurting Jamaica's competitiveness in international market by increasing domestic wages. Nguyen and Purnamasari (2011) argued that remittances might affect labor supply of migrant family members differently depending on both migrant's gender, and his or her influence on household decisions. Amuedo-Dorantes and Pozo (2006) stated that remittances help men to forego benefits of formal jobs and choose to do informal work. The absence of the husband because of migration would induce women with school-age children in remittance receiving families to leave the full time employment (Cabegin, 2006).

We look at another dimension in studying the effect of migration on the labor supply of migrants' family members in the source country -- their job satisfaction. Migrant's family members might consider remittances as their lost earning opportunities from not joining their migrant relatives in working abroad. A non-migrating member of a migrant's family would compare her own current earnings from working in the source country to what she might earn from migration basing on observed remittances from her migrant relatives. Additionally, a current or returning migrant provides information on existing labor market opportunities in the destination country. Using this information a non migrating member would create her own expectation on earnings from migration like if she joined the migrant, and compare them to her current wage. Then, the larger the difference between that what she receives now and her expected earnings from migration is, more dissatisfied from her current job she would be.

Individual expectations on earnings from migration might be affected by costs of migration, which are uncertain. Members of migrant families, however, have advantages in

reducing such costs basing on their migration experience. First of all, the cost of acquiring information on earnings possibilities in the destination country and on the job search would be lower for migrants' relatives because they learn this information from the migrants' experience. Migration costs are also lowered once the non-migrating members receive help from their migrant relatives in searching for jobs, housing and fulfilling all working and staying formalities in the destination country when they decide to migrate. Therefore, since the family is involved in migration, its members know how to reduce migration related costs, which allow them to get earnings almost close to their expected values.

Our discussion is consistent with the job satisfaction literature, which defines job satisfaction as an increasing function of the deviation of current workers' wages from the expected wages which they might receive from another employer or occupation. Introduced into the economics literature by Daniel Hamermesh (1977), in his economic model workers compare their wages in their current occupations with those from other job alternatives. If workers' current wages are higher than those from alternative jobs, they would be more satisfied with current jobs, and vice versa. His equilibrium condition at the time when an individual starts his work at the new occupation implies that there is no differential job satisfaction. Once the working experience with the current employer increases, the worker becomes more certain about her earning abilities that increase her job satisfaction. He finds a positive relationship between job satisfaction and the deviation of actual wages from the expected wages which are derived using information on the mean of the country's wage distribution conditioned on worker's individual characteristics such as experience, age, education and gender.

Hamermesh's findings have been confirmed across consequent studies. Clark and Oswald (1996) used two distinct variables in their regression analysis, logarithms of current and expected earnings, instead of a single variable of wage residuals. They found that while the coefficient on

the logarithm of current earnings is positive, the coefficient on the logarithm of expected income from other job alternatives is negative and statistically significantly different from zero. Comparison with alternative specifications allowed them to conclude that individual well-being does not depend on absolute income, but on the income comparison, i.e. on the relationship between what a person gets now and what she probably could get if she changed her job. In his following paper, Hamermesh (2001) finds that current shocks which widen earnings inequality also increases the current job satisfaction of those who are at the top of earning distribution. Diaz-Serrano and Vieira (2005) by analyzing European data found that the low-paid workers are less satisfied with their jobs compared to higher paid workers, except the British as they receive larger compensating non-pecuniary benefits. More recently, Card, Mas, Moretti, and Saez (2012) using a randomized manipulation of access to information among employees of the University of California find that granting access to earnings information of other employees increases job dissatisfaction among workers with wages less than the median in their pay unit and the same occupations.

In the next section, we discuss a simple model specification of job satisfaction, and incorporate migration into this model. In the third section we explain the semiparametric ordered response model, and discuss how we control for the endogeneity of migration related variables. The fourth section provides definitions and explanations of the data used in this paper. We used the data from 2007 World Bank Living Standard Measurement Survey on Tajikistan, a small Central Asian, former Soviet and transitional country which is highly dependent on migration and remittances. Differences in wages in Tajikistan and its migration destination country, Russia, along with increasing migration, make it a good country case for our study. The section five discusses estimation results of migration on the job satisfaction in Tajikistan. The final, fifth, section concludes.

### *1.2. Economic Models*

#### 1.2.1. Model of Job Satisfaction

We assume that an individual $i$ faces the following utility maximization problem with a constrained amount of leisure:

$$\max_{\{c\}} U^i\left(c^i, \bar{l}^i\right)$$

subject to the budget constraint:

$$w^i\left(T - \bar{l}^i\right) + v^i \geq \sum_{j=1}^{n} p_j c_j^i \,,$$

where $c_j^i$ is a consumption of good $j$ with corresponding price $p_j$, $c^i$ is individual $i$'s choice of the consumption bundle, $\bar{l}$ is the constrained amount of leisure, $T$ is the total available time, $w^i$ is the wage rate, and $v^i$ is a non-wage income. Assume also that standard conditions for the utility function along with Inada condition hold, i.e. $U_c > 0, U_{cc} < 0$, and $\lim_{c\to 0} U_c = \infty$, respectively.

Killingworth (1983) defines three main situations when such constrained leisure exists. Firstly, many firms for production efficiency set fixed hours of work and organize workers in several group-shifts. Then a person has the option either to take the job with the offered fixed hours of work or leave it. Secondly, person specific factors such as health issues might prevent workers from working more hours than some fixed number of hours. Finally, unemployment caused by imperfect information and imperfect mobility of people results in a discontinuous budget constraint. In such a case, individuals may not be able to immediately take up offers. This sets an upper limit to working hours per period, beyond which the budget becomes discontinuous. In all these situations, the income and substitution effects have little or no impact on individual

labor supply decisions. However, one important aspect of such model is that any possible increase in wages would result in increasing individual consumption. To see this we use Deaton and Muellbauer (1981) results on the linear function of the individual consumption with constrained labor supply.

Deaton and Muellbauer (1981) derived the following linear form of the restricted demand function for the consumption good $j$:[1]

$$\bar{c}_j^i = \left(\epsilon_j^i + \alpha_j^i \beta^i\right) + \left(\frac{\beta_j^i}{\beta^i} - \frac{\theta_j^i}{\theta^i}\right)\beta^i\left(\alpha^i - \bar{l}^i\right) + \left(\frac{\beta_j^i}{\beta^i} + \frac{\alpha_j^i}{\alpha^i - \bar{l}^i}\right)\left(w^i\left(T^i - \bar{l}^i\right) + v^i - \epsilon^i\right),$$

where $\alpha_j^i, \alpha^i, \beta_j^i, \beta^i, \epsilon_j^i, \theta_j^i$ and $\theta^i$ are preference parameters from individual $i$'s utility function.

Notice that an increase in the demand for the consumption good $j$ depends on wages, $w^i$, through total income: once the wage increases it would increase total income available for the individual in such way that she can spend more in buying the consumption goods. If we assume that there are no changes in individual non-wage income, then for any $w'^i > w^i$ the increment in consumption with constant labor supply can be defined as follows:

$$\bar{c}_j'^i - \bar{c}_j^i = \gamma_j^i\left(w'^i - w^i\right)\left(T - \bar{l}^i\right) > 0,$$

or, by summing over consumption goods $j$:

$$\bar{c}'^i - \bar{c}^i = \gamma^i\left(w'^i - w^i\right)\left(T - \bar{l}^i\right) > 0,$$

---

[1] See an equation (31) on the page of 1528.

[2] We, however, applied both parametric and semiparametric estimations on the sample with excluded zero wages, with the total number of observations of 2261. The estimate for *resm* in the parametric Model 2 is -0.1455 with standard error of 0.0691 (its marginal effect on job dissatisfaction is 0.0224). Its estimates in semiparametric Model 2 and Model 2-IV are -4.1198 with standard error of 5.7459 (its marginal effect on

where $\bar{c}'^{i} = \sum_{j=1}^{n} \bar{c}_j'^{i}$ , $\bar{c}^{i} = \sum_{j=1}^{n} \bar{c}_j^{i}$ , $\gamma_j^i = \frac{\beta_j^i}{\beta^i} + \frac{\alpha_j^i}{\alpha^i - \bar{l}^i}$, and $\gamma^i = \sum_{j=1}^{n} \gamma_j^i > 0$. These expressions are strictly positive since the demand for each consumption good is increasing in wage.

The last expression shows that the relationship between the expected individual demand for a bundle with more consumption goods and her current consumption can be expressed as the difference in work earnings: if a person wants to increase her consumption, such an increase should be compensated by receiving higher wages.

Next, using the mean value theorem, for any $t^i \in (0,1)$, we rewrite the difference between the individual utilities evaluated at $\bar{c}'^{i}$ and $\bar{c}^{i}$ in the following form:

$$\begin{aligned} U^i(\bar{c}^i, \bar{l}^i) - U^i(\bar{c}'^{i}, \bar{l}^i) &= -U_c^i\left(t^i\bar{c}^i - (1-t^i)\bar{c}'^{i}, \bar{l}^i\right)(\bar{c}'^{i} - \bar{c}^i) < 0 \\ &= -\rho^i(w'^{i} - w^i)(T - \bar{l}^i) < 0\,, \end{aligned} \tag{1.1}$$

which is strictly negative due to the imposed condition on utility function, $U_c > 0$, and $w'^{i} > w^i$, implying $\Delta\bar{c} > 0$; and $\rho^i$ is a random parameter driven by individual utility parameters for expected wage $w'^{i}$,

$$\rho^i = \gamma^i U_c^i\left(t^i\bar{c}^i - (1-t^i)\bar{c}'^{i}, \bar{l}^i\right) > 0\,. \tag{1.2}$$

The important result from (1.1) is that the comparison of utilities received from consuming different amounts of consumption goods could be made based on the difference between wages. The economic interpretation of this result is that, using the available information on the within source country wage distribution, an individual would construct her wage expectation from other possible job alternatives. In such a way she can evaluate the possible changes in her consumption if she decides to quit her current job in favor of new jobs with

different wages. If her wage expectation from outside jobs is higher than her current wage, or equivalently $w'^{i} > w^{i}$, then she would be unhappy with her current job. The outside wage is evaluated using the country's internal wage distribution:

$$w'^{i} = \int_{\min w_A}^{\infty} w dF_A(w) ,$$

where $\min w_A$ and $F_A(w)$ are the minimal wage rate and the wage distribution in the country $A$, respectively.

Notice that in the regression analysis, the expression (1.1) can be referred to as a Random Coefficient Model, since the parameter $\rho^{i}$ is random over population. The most useful way is to write $\rho^{i} = \bar{\rho} + u^{i}$ with $E(\rho) = \bar{\rho}$ and $E(u) = 0$, then expression (1.1) can be rewritten as:

$$\rho^{i}\left(w'^{i} - w^{i}\right)\left(T - \bar{l}^{i}\right) = \bar{\rho}\left(w'^{i} - w^{i}\right)\left(T - \bar{l}^{i}\right) + \tilde{u}^{i} ,$$

where $\tilde{u}^{i} = u^{i}\left(w'^{i} - w^{i}\right)\left(T - \bar{l}^{i}\right)$. The final expression has a constant coefficient on the wage differences which is a parameter of interest, as well as the interaction term between the unobserved heterogeneity and wage differences. Therefore, one also needs to calculate the average partial effects of model variables, by averaging over unobserved $\tilde{u}^{i}$.

Using the last expression we can rewrite the function of job satisfaction. First, notice that an individual would be satisfied if her current wage is greater than that which she might receive from any other employer: $U^{i}\left(\bar{c}^{i}, \bar{l}^{i}\right) - U^{i}\left(\bar{c}'^{i}, \bar{l}^{i}\right) > 0$. Therefore, an individual would compare her current wage, $w^{i}$, to the possible wage that she could receive in another job, $w'^{i}$, for the same hours of work based on her individual worker characteristics and current market conditions.

Define by $J$ the index of individual $i$'s job satisfaction; for unknown cut points $\tau_1 < \tau_2 < \cdots < \tau_{k-1}$:

$$\begin{aligned} J_1 = 1: &\quad \alpha_w\left(w'^i - w^i\right)\left(T - \bar{l}^i\right) + \tilde{u}^i \le \tau_1 \\ J_2 = 2: &\quad \tau_1 < \alpha_w\left(w'^i - w^i\right)\left(T - \bar{l}^i\right) + \tilde{u}^i \le \tau_2 \\ \vdots &\quad \vdots \\ J_k = k: &\quad \tau_{k-1} < \alpha_w\left(w'^i - w^i\right)\left(T - \bar{l}^i\right) + \tilde{u}^i \end{aligned} \tag{1.3}$$

Using this specification we can estimate the effect of the difference between the individual expected wages from other jobs and current wages on the job satisfaction. Since by construction $\rho^i > 0$, current job satisfaction from having lower wages, i.e. increasing wage difference of $\left(w'^i - w^i\right)$, would be a simple t-test on the negative sign of the coefficient $\alpha_w$.

### 1.2.2. Model of Migration and Job Satisfaction

Once a household sends a migrant, its members acquire information about outside country wage distribution through either the size of remittances, or information directly received from a migrant. Having such information, a member $i$ of the migrant's family would construct her expectation on her earnings from migration as if she has migrated:

$$r'^i = \int_{\min w_B}^{\infty} w dF_B(w)\,,$$

where $\min w_B$ and $F_B(w)$ are the minimal wage rate and the wage distribution in the destination country $B$, respectively.

Therefore, with such information she would be able to compare her utility based on her earnings in her source country with her utility from her expected earnings in the destination country:

$$U^i(\bar{c}^i, \bar{l}^i) - U^i(\bar{c}^{r\,i}, \bar{l}^i) = -\rho^{i,r}(r^i - w^i)(T - \bar{l}^i) = -(\bar{\rho}^r(r^i - w^i)(T - \bar{l}^i) + \tilde{u}^{i,r}), \quad (1.4)$$

where $r^i$ is the expected wage earnings from migration by individual $i$, $\bar{c}^{r\,i}$ is an individual $i$'s target consumption if she migrated, $\tilde{u}^{i,r} = u^{i,r}(r^i - w^i)(T - \bar{l}^i)$ is an heteroscedastic error term, and $\bar{\rho}^r$ with $\rho^{i,r}$ reflect changes in utility parameters with observed remittances.

Her current job satisfaction depends on two parallel utility comparisons defined as in (1.1) and (1.4):

$$[U^i(\bar{c}^i, \bar{l}^i) - U^i(\bar{c}'^i, \bar{l}^i)] + [U^i(\bar{c}^i, \bar{l}^i) - U^i(\bar{c}^{r\,i}, \bar{l}^i)]$$
$$= -(\bar{\rho}(w'^i - w^i)(T - \bar{l}^i) + \bar{\rho}^r(r^i - w^i)(T - \bar{l}^i) + \tilde{u}^i + \tilde{u}^{i,r})$$

We estimate this equation for migrant family members working in the source country. Notice that we intentionally add both wage differences in the equation, which allow us to estimate the effect of the difference of the expected outside country wages and current wages of migrants' relatives on their job satisfaction keeping constant the difference in the expected internal country wage and current wages.

Starting from this point, we distinguish these two differences by calling the first difference, i.e. the difference between individual expected wages from internal country jobs and individual's current wages, as the *intra-country* wage difference. We call the second difference, i.e. the difference between individual expected wages from migration (or the destination country wage distribution) and individual's current wages, as the *inter-country* wage difference.

One can also interpret the last equation using the definition of first order stochastic dominance. If the destination country's wage distribution dominates the wage distribution in the source country in the sense of the first stochastic dominance, expected utility from migration

would be higher than the expected utility of changing jobs within the source country: $F_B(w) \leq F_A(w) <=> U^i(\bar{c'}^i, \bar{l}^i) \leq U^i(\bar{c}^{ri}, \bar{l}^i)$. This implies a significant negative effect of the inter-country wage difference, $(r^i - w^i)$, on the their job satisfaction. A similar argument works, if one apply the second order stochastic dominance in considering the wage distributions between the source and destination countries for certain occupations or workers' other individual characteristics.

Using this expression we can rewrite the job satisfaction index function (1.4) including information on inter-country wage differences of the migrants' relatives:

$$\begin{aligned} J_1 = 1: &\quad \alpha_w(w'^i - w^i)(T - \bar{l}^i) + \alpha_r m^{i,h}(r^i - w^i)(T - \bar{l}^i) + \tilde{\zeta}^i \leq \tau_1 \\ J_2 = 2: &\quad \tau_1 < \alpha_w(w'^i - w^i)(T - \bar{l}^i) + \alpha_r m^{i,h}(r^i - w^i)(T - \bar{l}^i) + \tilde{\zeta}^i \leq \tau_2 \\ \vdots &\quad \vdots \\ J_k = k: &\quad \tau_{k-1} < \alpha_w(w'^i - w^i)(T - \bar{l}^i) + \alpha_r m^{i,h}(r^i - w^i)(T - \bar{l}^i) + \tilde{\zeta}^i \ , \end{aligned} \tag{1.5}$$

where $m^{i,h} = 1$ if the households $h$ of the individual $i$ has any migrant, and, $\tilde{\zeta}^i = -\tilde{u}^i - m^{i,h}\tilde{u}^{i,r}$ is a composite heteroscedastic error term. According to our discussions above, $\alpha_w$ and $\alpha_r$ both should have negative signs.

### *1.3. Econometric Model*

#### 1.3.1. Semiparametric Ordered Response Model

Both models in (1.3) and (1.5) imply heteroscedastic error terms $\tilde{u}^i$ and $\tilde{u}^{i,r}$. The estimation of such models using standard parametric ordered response models could be problematic. According to Wooldridge (2010), the current concerns in parametric estimation are mainly about the signs of the model coefficients as well as their magnitudes. Firstly, if parametric response models are applied, the heteroscedastic error terms might affect the signs of partial effects of the model

variables in such way that the true coefficients of model variables would have different signs from the partial effects of those variables. Secondly, in parametric ordered models the signs of estimated coefficients do not necessarily determine the directions of corresponding variable effects on model intermediate outcomes (i.e. for $m = 2, \dots, k-1$), because of symmetry and monotonicity properties the standard normal probability distribution function, as well as the size of the cut points. And, finally, the parametric estimation of response models with endogenous variables would produce scaled estimates, thus to derive the original values of coefficients can be estimated by dividing them by bootstrapped standard errors, or using the delta method.

We use the semiparametric estimation for models (1.3) and (1.5), which is based on results from Klein and Spady (1993), Blundell and Powell (2004), and Rothe (2009). The main advantage of semiparametric methods in estimating our job satisfaction model is that it allows us to relax the distributional assumptions on the error terms of the model $\tilde{u}^i$ and $\tilde{u}^{i,r}$. Such advantage is crucial, since in the parametric model the consistency of estimators is sensitive to the distributional assumption of the error term (Klein & Sherman, 2002).

Firstly, we impose the single index restriction for probabilities of outcomes of reported job satisfactions $Y_m = 1$, as in categorical numbers of $m = 1,2 \dots k$, conditional on data $X_1, X_2, X_3$ by

$$\begin{aligned} \mathrm{E}(Y_m = 1 | X_1, X_2, X_3) &= \Pr(X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + C_0 + u \geq 0) \\ &= \Pr(\beta_1(X_1 + X_2\theta_1 + X_3\theta_2) + C_0 + u \geq 0) \\ &= \Pr(X_1 + X_2\theta_1 + X_3\theta_2 + \epsilon \geq 0) = \Pr(V_1 \geq -\epsilon) = F_m(V_1)\,, \end{aligned}$$

where $Y_m = 1\{J_m = k\}$, $\beta$'s are original coefficients of the model, $\theta$'s are ratios of original coefficients to $\beta_1$, $C_o$ is the constant of the model, $V_1 = X_1 + X_2\theta_1 + X_3\theta_2$ is an index, $u$ is an error term with $E(u) = 0$, and $\epsilon = C_0 + u$, $F_m$ is the cumulative density function. For

identification and consistency purposes $X_1$ should be a continuous variable and there should be no other functions of $X_1$ in the model, a matrix $(X_1, X_2, X_3, \mathbf{1})$ has a full rank as $N \to \infty$.

Such restrictions allow us to improve the finite sample behavior of our estimator by keeping the dimension of the data small, to apply estimation even when the index has a non-linearly functional form. This restriction allows estimating a ratio of coefficients ignoring the constant term along with the thresholds. Imposing such restriction, however, do not help us to recover the original coefficients of the model.

Using such index restriction, we would be able to derive the conditional distribution of $Y_m$ on model's data $X_1, X_2, X_3$ using the following conditional expectations:

$$\begin{aligned} &\mathrm{E}(Y_1 = 1|X_1, X_2, X_3) = F_1(V_1) \\ &\mathrm{E}(Y_2 = 1|X_1, X_2, X_3) = F_2(V_1) \\ &\vdots \\ &\mathrm{E}(Y_m = 1|X_1, X_2, X_3) = F_m(V_1) \end{aligned}$$

Each expectation could be derived using a single-index binary model discussed in Klein and Spady (1993). Hence, using the above probabilities we can write the quasi loglikelihood function in the following form:

$$\ln \mathcal{L} = \frac{1}{N} \sum_i \sum_m \pi_i \ln\big(F_m(V_{1i})\big) ,$$

where $\pi_i$ is a trimming function, which helps to keep the probabilities away from the end of tails, and $N$ is a sample size.

These probabilities can be estimated using the kernel regression estimator

$$\hat{F}_m(\hat{V}_{1i}) = \frac{\sum_{j\neq i} Y_{m,i} K\left(\frac{\hat{V}_{1i}-\hat{V}_{1j}}{h_1}\right)\Big/\left((N-1)h_1\right)}{\sum_{j\neq i} K\left(\frac{\hat{V}_{1i}-\hat{V}_{1j}}{h_1}\right)\Big/\left((N-1)h_1\right)},$$

where $K$ is a Gaussian kernel function, and the bandwidth $h_1 = \sigma_{\hat{V}_1} N^{-1/5}$ and $\sigma_{\hat{V}_1}$ is a standard deviation of $\hat{V}_1$(see Silverman (1986)).

### 1.3.2. Semiparametric Estimation with Endogenous Explanatory Variable

The main problem in estimating the effects of migration are the endogeneity issue, as both the decision on emigration, and, consequently, the receipt of remittances are not random events. Households are self-selected in sending their member(s) abroad; as well migrants are self-selective in returning to their home countries. In addition to these emigration self-selection issues, the duration depending heterogeneity, i.e. the decision on when to migrate, could cause the biasness in estimators (Gibson, McKenzie, & Stillman, 2010). The selection issue of migration in our model on the job satisfaction and the migration relationship arises as only working migrant relatives in the source country can compare their current work earnings to those from migration, while workers who do not have migrant relatives cannot. Endogeneity problems also rise when there is a simultaneity issue between an individual's job dissatisfaction and migration of the family member. Close relatives of individuals who are mostly dissatisfied with wages they receive at their current jobs due to their altruistic preferences might choose to migrate and consequently to send remittances in order to help in filling this person's needs. In such way, the coefficient on differences of wage and remittances for families with current migrants might be upward biased. Modeling unobservable variables like unreported income other than wages (such as income from informal employment) can also influence individual job satisfaction. Estimating

the model without controlling for such income would produce downward biased estimates on the difference of remittances and individual current wages.

There are several ways to deal with endogeneity issues. The most popular is the instrumental variable approach. To apply similar to the instrumental variable approach to our regression analysis in respect to the inter-country wage difference, we refer to results of Blundell and Powell (2004) and Rothe (2009). They developed a semiparametric method for estimating binary response models with continuous endogenous regressor, which can be extended to semiparametric ordered response model. However, since the endogenous explanatory variable in our model has a truncated distribution (i.e. we have do not observe outside wage differences for members in non-migrant families), instead of using the ordinary least squares estimation for the first stage reduced form equation, we use Ichimura's semiparametric non-linear least squares (1993).

We specify our semiparametric ordered response model with an endogenous explanatory variable:

$$\mathrm{E}(Y_m = 1|X_1, X_2, X_3^e) = \Pr(X_1 + X_2\theta_1 + X_3^e\theta_2 \geq -\epsilon^e) = \Pr(Y_m = 1|V_1^e \geq -\epsilon^e)\,, \qquad (1.6)$$

where one of explanatory variables, $X_3^e$, is endogenous, and superscript $e$ in index $V_1$ implies that it has an endogenous variable as its argument.

The endogenous variable $X_3^e$ is assumed to be determined by the reduced non-linear form:

$$X_3^e = \psi(Z + X_1\alpha_1 + X_2\alpha_2) + \varepsilon\,,$$

where $\varepsilon$ is a stochastic error term, $\psi(\cdot)$ is an unknown function, $\alpha$ are coefficients normalized by the coefficient of excluded from the structural equation, a continuous variable of $Z$, $(Z, X_1, X_2)$ is a matrix of all exogenous variables, which has a full rank with probability 1. Then by construction we would have:

$$E(\varepsilon|X_1, X_2, Z) = 0\,.$$

By defining an index $V_2 = Z + X_1\alpha_1 + X_2\alpha_2$, we can rewrite the conditional expectation of outcome $Y_k$ as:

$$\mathrm{E}(Y_m = 1|X_1, X_2, X_3^e, V_2) = \Pr(V_1^e \geq -\epsilon^e|V_2) = F_m(V_1^e, V_2)\,,$$

where $F_m(\cdot)$ is a cumulative distribution function of $Y_m = 1$ conditioned on two indexes, $V_1^e$ and $V_2$. Therefore, the semiparametric ordered response model with a continuous endogenous explanatory variable can be characterized as a double index model.

We rewrite the quasi log likelihood function in the following form:

$$\ln \mathcal{L} = \frac{1}{N}\sum_i \sum_m \pi_{i,X_{1i}}\pi_{i,V_{2i}} \ln\big(F_m(V_{1i}, V_{2i})\big)\,,$$

where $\pi_X$ and $\pi_{V_2}$ are trimming functions on continuous variables in $(X_1, X_2, X_3^e)$, and $V_2$, respectively.

$V_2$ is estimated in the first stage by running the Semiparametric Nonlinear Least Squares of $X_3^e$ on $(Z, X_1, X_2)$. Then conditioning on the estimates of the first stage index $\hat{V}_2$, we can estimate functions $F_k\big(V_{1i}, \hat{V}_{2i}\big)$ by the kernel regression estimator:

$$\hat{F}_m(\hat{V}_{1i}, \hat{V}_{2i}) = \frac{\sum_{j \neq i} Y_{mi} K\left(\frac{\hat{V}_{1i} - \hat{V}_{1j}}{h_1}\right) K\left(\frac{\hat{V}_{2i} - \hat{V}_{2j}}{h_2}\right) \Big/ \left((N-1)h_1 h_2\right)}{\sum_{j \neq i} K\left(\frac{\hat{V}_{1i} - \hat{V}_{1j}}{h_1}\right) K\left(\frac{\hat{V}_{2i} - \hat{V}_{2j}}{h_2}\right) \Big/ \left((N-1)h_1 h_2\right)}.$$

The bandwidth for two-index model is chosen as $h_1 = \sigma_{\hat{V}_1} N^{-1/6}$ and $h_2 = \sigma_{\hat{V}_2} N^{-1/6}$, where $\sigma_{\hat{V}_1}$ and $\sigma_{\hat{V}_2}$ are standard deviations of $\hat{V}_1$ and $\hat{V}_2$, respectively.

To confirm that this semiparametric estimation method performs well, we decided to fulfill the experiment using Monte Carlo simulations. The experimental data was generated using the following similar structure of our model:

$$I = 1\{2X_1 + X_2 - X_3 - X_4 - 3 > 0\},$$

$$Y_2 = \begin{cases} 2X_1 + X_2 - X_3 - X_4 - 3 + \varepsilon & : I = 1 \\ 0 & : I = 0 \end{cases}$$

$$Y_1^* = 3X_2 - X_3 + 2X_4 + 3Y_2 + 1 + \epsilon \ : \ \epsilon = 5^{-1/2}(2\varepsilon + u),$$

$$\begin{array}{lll} Y_1 = 1 & \text{if} & Y_1^* \leq \tau_1 \\ Y_1 = 2 & \text{if} & \tau_1 < Y_1^* \leq \tau_2 \\ Y_1 = 3 & \text{if} & Y_1^* > \tau_2 \ , \end{array}$$

where all $X$'s, $\varepsilon$ and $u$ have independent normal distributions. Cut points $\tau$'s are defined using tertiles of $Y_1^*$. The sample size is 3000; the number of Monte Carlo replications is 1000.

It is easily seen from these equations that the true semiparametric coefficients in the reduced form equation for $Y_2$ are $(0.5, -0.5, -0.5)'$, and in the structural equation for $Y_1$ are $(-0.33, 0.67, 1)'$. The mean, median and standard deviation of distributions of parameter estimates of the reduced form equation for $Y_2$ from Monte Carlo simulations are $(0.5012, -0.5006, -0.5003)'$, $(0.5011, -0.5002, 0.4990)'$ and $(0.0315, 0.0297, 0.0313)'$,

respectively. The mean, the median and standard deviations of distributions of parameter estimates of the structural equation for $Y_1$ are $(-0.3265, 0.6793, 0.9590)'$, $(-0.3262, 0.6791, 0.9587)'$ and $(0.0132, 0.0166, 0.0507)'$, respectively. Our Monte Carlo experiment shows that the coefficient estimates from both reduced form and structural equations estimated using the suggested semiparametric estimation methods, are very close to their true values, they have both negligible biases and smaller variances.

### *1.4. Empirical study*

#### 1.4.1. Tajikistan's Case

We have chosen the country case of Tajikistan for several reasons. Firstly, it is a transitional country which currently experiences an increasing labor migration due to high wage differences between Tajikistan and the main destination of its migrants, Russia. The average real wages in the Russian Federation in 2010 were about 8.5 times larger than those in Tajikistan. Such wage differences not only drive more people from Tajikistan to Russia, but might also increase the dissatisfaction among current workers in Tajikistan with their current wages.

Secondly, Tajikistan and Russia share the 70 year history of association in a single country, the USSR, under similar identities, cultural norms and traditions, where people use Russian as an international communication language. Such a commonly shared historical background helps to lower migration costs. Some elder generation of Tajiks still speak Russian and hold diplomas from Soviet schools and universities, which are helpful in finding jobs in Russia. They also do not need to spend additional time and money in learning Russian language. Some Tajikistan's migrants might rely on help from their older Russian friends and families, whom with they used to work, or served in the Soviet army, in finding jobs and temporary accommodations.

Furthermore, since families heavily depend on remittances and because of migration being seasonal in Tajikistan (where migrants return in each winter after seasonal job cuts due to the Russia's cold weather), migrant families observe perfectly the wages of their migrant relatives through either the remittances they receive or directly from migrants themselves. The World Bank reports that the size of remittances sent by Tajik migrants reached one-third of the country's GDP in 2009 ranking Tajikistan as the world's most highly dependent country on remittances (World Bank, 2011). According to 2007 World Bank Living Standards Measurement Survey on Tajikistan, about 27.35% of interviewed households received remittances in last 12 months, and 29% of those households which received remittances are heavily dependent on them. Furthermore, the International Labor Organization reports that 77% of returned Tajik migrants confirmed that they plan to migrate again in the next working season (International Labor Organization, 2010). Such seasonality, easy and accurate observance of migrant's earnings would help other migrant family members, who remained in Tajikistan to build their expectations on their possible earnings from migration if they joined their migrant relatives.

International migration is relatively new phenomena in Tajikistan. As a country-member of the former Soviet Union, international migration was strictly controlled and even "prohibited" by the Central Soviet Government. After the Union's collapse, this restriction was removed, thereby involving an appreciably large proportion of Tajikistan's population.

These initial conditions make Tajikistan a good case to study, where one does not need to be very concerned about historically well-established patterns and traditions of migration and allowing us to focus only on economic issues and factors which help to explain how these two processes interact. Tajikistan's current migration experience and features allow us to examine our theoretical model in studying the effect of migration on the job satisfaction of migrant family members.

### 1.4.2. Data and Variable Definitions

As a part of the response to the recognition of current migration trends in Tajikistan, data was collected in 2007 World Bank Living Standard Measurement Survey (2007 WB LSMS) highlighting migrants and their families. This survey includes questions on migration, education, health, labor market, housing, transfers and social assistance, subjective poverty and food security, as well as data for household's expenditure and income. There were 4860 households surveyed in Tajikistan in 2007, 745 households had current migrants, and there were 982 migrants in total.

We look at the reported overall individual satisfaction from current primary jobs in Tajikistan. The survey asks a question "Overall how satisfied are with your job?". The answers are recorded for those who were present in the household during the survey as "Very satisfied", "Satisfied", "Neither satisfied nor dissatisfied", "Dissatisfied", and "Very dissatisfied". Because few observations were reported at extreme values, we put two first answer categories ("Very satisfied" and "Satisfied") together into one category and named it "Satisfied", and two last categories ("Dissatisfied" and "Very dissatisfied") into another single category, and named it "Dissatisfied". This categorical variable is used as the dependent variable in our regression analysis.

The sample size is 3022, including individuals with zero reported wages. We have not excluded them for two reasons. Firstly, working individuals, who reported their job satisfaction, work at different employers that include family owned businesses and farms. In such businesses and farms, involved family members do not necessarily receive individual wages in cash (i.e. they have zero reported individual work earnings), since they work at increasing family's total income which is common. Secondly, since employment in the informal sector is common in Tajikistan,

many families have other than wages income from employment that might be not reported (Abdulloev, Gang, & Landon-Lane, 2012). Since, we would allow non pecuniary effects and other non-reported income from current employment to be a part of the error term in our model; we did not exclude these observations from the sample.[2]

We are interested in estimating the effects of two variables on job satisfaction; that is the effect of intra-country and inter-country wage differences.[3] The variable intra-country wage difference is constructed as the difference between the reported work earnings, which includes cash, bonuses and in-kind payments, and the expected value of work earnings from the country's internal wage distribution, which are calculated using Mincer's (1970)‘s earnings regression equation for each provinces of Tajikistan with division into rural and urban areas (totally 9 geographical areas). The variable inter-country wage difference is constructed as the difference between reported work earnings, which includes cash, bonuses and in-kind payments received by non-migrating members of migrant's families in Tajikistan, and, the expected value of work earnings using parameters of the estimated Mincer's earnings regression equation for current migrants. The value of the variable on the inter-country wage differences for working individuals in non-migrant families is zero, since according to our economic model, they cannot observe information on wage distributions of destination countries (by other words, the spillover effects of migration are set at zeros). This selection issue is accounted for in our semiparametric Model 2-

[2] We, however, applied both parametric and semiparametric estimations on the sample with excluded zero wages, with the total number of observations of 2261. The estimate for *resm* in the parametric Model 2 is -0.1455 with standard error of 0.0691 (its marginal effect on job dissatisfaction is 0.0224). Its estimates in semiparametric Model 2 and Model 2-IV are -4.1198 with standard error of 5.7459 (its marginal effect on job dissatisfaction is 0.0039), and -19.1018 with standard error of 6.7299 (its marginal effect on job dissatisfaction is 0.0205), respectively.

[3] The appropriate term should be "salary" instead of "wage", because monthly salaries were recorded in the data. We, however, choose to stay with the "wage" term in order to avoid confusion in the discussion of the previous sections of this paper.

IV (see the next section). Variables which were included in Mincer's earning regression equation include individual age, gender, and levels of education.

Other exogenous variables in the model of job satisfaction include dichotomous variables defining whether an individual has the highest level of education from any technical school (vocational education), whether the individual has the highest level of education from the university, whether an individual is male, whether an individual lives in the capital, whether the job is affiliated with a social security scheme (i.e. the National Social Protection Fund that is used to cover expenses on social protection of employees), whether the working place is in a fixed building, whether an individual works in the street or market. The model also includes continuous explanatory variables on individual ages, the number of children in the household, and the total value of durable goods owned by families as a proxy for non-wage income. Definitions of these variables are provided in Table1.1.

As it was briefly discussed above, a main problem in estimating effects of migration is that the migration related variable in our model – the variable on the inter-country wage difference – is endogenous. There are several ways to deal with endogeneity issue, but the most popular is the instrumental variable approach. Instrument variables, however, vary depending on the subject of studies. Brown and Leeves (2007) used migration networks to instrument the number of migrants in the household. This instrument is constructed using the community level migration patterns. McKenzie and Rapoport (2007) suggest instead using historic networks as an instrument for migration since communities are affected by external shocks that would lead to changes in current migration patterns. While migrant networks are widely used as an instrument to the decisions on family involvement into migration, there are other instrumental variables applied to migration such as distances to roads and main cities, and economic changes.

Since there was no migration history in Tajikistan as it was mentioned above, we used the current migration network per local communities as an excluded variable to control for endogeneity of migration in our semiparametric model. The migrant network variable at the community level is defined as a share of community's migrants in the total number of adults in that community (there are 269 communities in the sample). Adults are defined as those who are 16 years old and above. We define migration network per local community as,

$$netw_R = \frac{\sum_{R,h} m_{R,h}}{\sum_{R,h} n_{R,h}^{ad}},$$

where $netw$ is network variable defined for each community $R$, $m_{R,h}$ is a number of migrants in household $h$ in the community $R$, and, $n_{R,h}^{ad}$ is a number of adults in household $h$ in the country $R$. Since this variable is defined per community level it is exogenous to individual decisions.

Table 1.2 reports summary statistics of variables for three separate groups based on reported job satisfaction: dissatisfied, neither satisfied nor dissatisfied, and satisfied. There are 242 people who reported being dissatisfied from their jobs: 64 of them have migrant relatives and 178 people do not have. 668 people reported being neither satisfied nor dissatisfied from their jobs: 159 of them live in families with migrants, and 509 people live in families without migrants. A larger number of people, 2112, are in the group who reported being satisfied from their jobs, out of whom 457 people have migrant relatives, and 1655 people do not have migrant relatives.

Table 1.2 shows that individually reported job satisfaction increases with age (*age*). This result is consistent with findings of Hamermesh (1977), and is probably due to decreasing worker's uncertainty about her future wage distribution. Higher dissatisfaction at younger ages might imply that, firstly, the people do not develop job-specific human capital, consequently, they are less paid relatively to elder workers. With smaller wages, younger workers are more likely

being dissatisfied from their jobs than elder workers. It also might be because of the younger workers' mismatch with their current jobs. Since mismatch leads to lower wages, we can hypothesize again that younger workers exhibit higher dissatisfaction relatively to elder workers. There are also differences in age means between individuals living in families with and without migrants: those people who have migrant relatives are older than their cohorts in the same job satisfaction category. This is not surprising if one takes into account the fact that migrants in Tajikistan are predominantly young men, and because of their absence, the mean age of migrant family members increases.

Both the variables on inter- and intra-country wage differences (*resm* and *resw*) are increasing with the job dissatisfaction. Since these variables are constructed as difference between expected wages that individual could receive from other similar jobs either abroad or within the same province of Tajikistan, and her current wages, the increase in these variables would imply that the individual receives less than an average person with similar age and educational background does. The larger these gaps, the more dissatisfied people would be with their jobs because of being underpaid. Another interesting picture is that the distribution between these two variables, the inter- and intra-country wage differences, which are significantly different: the variable on inter-country wage difference is larger on mean than the intra-country wage difference. This difference is due to lower wage distribution in Tajikistan compared to migrants' earnings in their main destination country, Russia.

Number of children (*ch14*) in the family does not show any monotonic relationship with job satisfaction. Individuals living in families with relatively more children have reported at average being neither satisfied nor dissatisfied with their jobs. The amount of durable goods (*durs*) owned by families do not significantly differ among groups with reported dissatisfaction

and neither satisfaction nor dissatisfaction. However, the satisfied group of people have a higher amount of family owned durable goods.

The level of education from vocational schools (*meduc*) does not differ among job satisfaction groups, but differs between individuals living in migrant and non-migrant families. The educational level from universities (*heduc*) does not differ among individuals living in families with migrants and without migrants in the dissatisfied group. However, the gap in shares of people with university education between those who live in families with migrants and without migrants increases with satisfaction, as their averages over satisfaction groups do: for neither satisfied nor dissatisfies group, the difference in shares of people with university degree between migrants' relatives and without is about 4.7%, at average 9.2% of people in this group have a degree at least from universities; while, these numbers are 6.1% and 17% are for satisfied group, respectively. This observation is consistent with the fact that the families are self-selected into migration: Tajikistan's families with members with lower skills or lower levels of education chose to be involved into migration, while people with higher education, or professionals, have more opportunities to engage in "unreported" income from their formal jobs, and prefer to remain in Tajikistan (Abdulloev, Gang, & Landon-Lane, 2012). Such access to "unreported" income by professionals might be a reason for their satisfaction from current jobs.

The gender variable (*male*) also differs between working individuals in families with and without migrants over groups of reported job satisfaction. Individuals who reported being satisfied from their current jobs are 57.1% are men living in non-migrant families and 42.9% are men living in families with migrants. Among those who reported being neither satisfied nor dissatisfied from their jobs are 51.5% live in families without migrants and 49.1% live in families with migrants. In the dissatisfied group, 50% of those who live in families without migrants are men, while 43.7% of working migrants' family members are men. This large difference between

working members of families without and with migrants is due to male dominance in migration in Tajikistan. A larger number of people living in the capital of Tajikistan (*capl*) have reported being neither satisfied nor dissatisfied from their current jobs. The share of people living in the capital city is smaller for families with migrants over all three job satisfaction groups which indicates that Tajikistan's migrants are predominantly from rural areas.

Affiliation with the social security scheme (*ssec*) of employers increases monotonically with reported job satisfaction. People working at employers who are affiliated with social security scheme feel more "secure" about their future, post retirement pension, and receive state health benefits in cases of emergencies. The social security affiliation might also imply that workers have long term contracts with their employers, as well as employers are being well-established companies, which increases individual job satisfaction A share of people who work in fixed premises (*fdpl*), such as offices or plants, in average increases over job satisfaction groups. Conversely, a share of people who work in the street or markets (*smpl*) decreases over job satisfaction groups. Such different relationships between these two workplaces and the job satisfaction might be explained to the fact that the work within fixed buildings and premises is affiliated with the social security scheme, long term contracts and well-established employers, while working on streets and markets implies self-employment with an absence of social security, or at small and "young" companies.

The last variable in our list is the community level migrant networks (*netw*). This variable does not significantly differ across job satisfaction groups: the mean of the migrant network variable for dissatisfied group is 0.0895, for a neither satisfied nor dissatisfied group is 0.0849, and for a satisfied group is 0.0865. The variable's mean, however, significantly varies between individuals living in families with and without migrants: for the dissatisfied group, the mean of the migrant network for migrant relatives is 0.1101, while, for people without migrant relatives, it

is 0.0689; for the neither satisfied nor dissatisfied group, the mean of the variable for migrant relatives is 0.1007, but it is lower again for people without migrant relatives and equals to 0.0692; finally, for the satisfied group, the corresponding means of network variable for people with and without migrant relatives are 0.1026 and 0.0705, respectively. Such non-variation of the migrant network variable across job satisfaction groups, and its variation between people living in families with and without migrants, makes it a valid instrument for migration related variable of our model.

In the next section, we use multivariate regression analysis in order to specify the partial marginal effects of migration on individual job satisfaction.

**Table 1.1. Job Satisfaction Model Variable Descriptions**

| Variables | Descriptions |
|---|---|
| age | Individual's age. |
| resm | Difference between expected wages from migration and current work earnings, in thousands of Somoni. |
| resw | Difference between expected intra-country wages and current work earnings, in thousands of Somoni. |
| ch14 | Number of children in families with age less than 15. |
| durs | Current value of durable goods owned by families, in thousands of Somoni. |
| meduc | Dummy variable on whether an individual holds the highest level of education from the vocational school. |
| heduc | Dummy variable on whether an individual holds the highest level of education from university. |
| male | Dummy variable on whether an individual is male. |
| capl | Dummy variable on whether an individual lives in the capital city (Dushanbe). |
| ssec | Dummy variable on whether an individual's job is affiliated with social security scheme. |
| fdpl | Dummy variable on whether an individual workplace is in a fixed building. |
| smpl | Dummy variable on whether an individual workplace is in the street or market. |
| netw | Network variable (excluded exogenous continuous variable). |

**Table 1.2. Summary Statistics of Job Satisfaction Model Variables**

| Variable | Dissatisfied | | | | Neither Satisfied Nor Dissatisfied | | | | Satisfied | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | With migrant | | Without migrant | | With migrant | | Without migrant | | With migrant | | Without migrant | |
| | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. |
| age | 38.7656 | 13.1218 | 37.9157 | 11.9158 | 40.1509 | 13.6044 | 39.8723 | 12.3396 | 42.2101 | 12.8121 | 40.6073 | 11.9257 |
| resw | 0.2536 | 0.2499 | 0.1942 | 0.2397 | 0.2216 | 0.2733 | 0.2125 | 0.2552 | 0.1954 | 0.2690 | 0.1602 | 0.4242 |
| resm | 1.1819 | 0.2798 | 0.0000 | 0.0000 | 1.1813 | 0.3118 | 0.0000 | 0.0000 | 1.1744 | 0.3105 | 0.0000 | 0.0000 |
| meduc | 0.1719 | 0.3803 | 0.1966 | 0.3986 | 0.1572 | 0.3652 | 0.1690 | 0.3751 | 0.1751 | 0.3804 | 0.2042 | 0.4033 |
| heduc | 0.0938 | 0.2938 | 0.0899 | 0.2868 | 0.0692 | 0.2546 | 0.1159 | 0.3204 | 0.1400 | 0.3474 | 0.2012 | 0.4010 |
| male | 0.4375 | 0.5000 | 0.5000 | 0.5014 | 0.4906 | 0.5015 | 0.5147 | 0.5003 | 0.4289 | 0.4955 | 0.5710 | 0.4951 |
| ch14 | 2.0938 | 1.5808 | 2.3146 | 1.6847 | 2.7107 | 1.8154 | 2.5973 | 1.9784 | 2.3217 | 1.8471 | 2.2719 | 1.6960 |
| capl | 0.0781 | 0.2705 | 0.0899 | 0.2868 | 0.0629 | 0.2435 | 0.1729 | 0.3785 | 0.0525 | 0.2233 | 0.1057 | 0.3076 |
| ssec | 0.3750 | 0.4880 | 0.3258 | 0.4700 | 0.4151 | 0.4943 | 0.3536 | 0.4786 | 0.4333 | 0.4961 | 0.4792 | 0.4997 |
| fdpl | 0.2813 | 0.4532 | 0.2247 | 0.4186 | 0.2516 | 0.4353 | 0.2849 | 0.4518 | 0.3326 | 0.4717 | 0.4048 | 0.4910 |
| smpl | 0.1406 | 0.3504 | 0.2022 | 0.4028 | 0.0943 | 0.2932 | 0.2083 | 0.4065 | 0.1007 | 0.3012 | 0.1184 | 0.3232 |
| durs | 1.8090 | 3.0430 | 2.7754 | 5.0099 | 1.7832 | 3.9604 | 2.2341 | 6.0233 | 3.4087 | 7.9170 | 3.4327 | 9.0020 |
| netw | 0.1101 | 0.0574 | 0.0689 | 0.0461 | 0.1007 | 0.0571 | 0.0692 | 0.0479 | 0.1026 | 0.0509 | 0.0705 | 0.0468 |
| Observations | 64 | | 178 | | 159 | | 509 | | 457 | | 1655 | |

### 1.4.3. Regression Analysis

Tables 1.3 and 1.4 report the results from estimating the effect of an array of the variable specified above on individual job dissatisfaction using the parametric ordered probit and semiparametric ordered response models. There are two models are reported in each table: in the Model 1, we estimate the effects of all mentioned exogenous variables except the variable on the inter-country wage difference for individuals living in families with migrants; and, in the Model 2, we estimate the same model as in Model 1 but with the inclusion of the variable on the inter-country wage difference. Notice that we allow both intra- and inter-country wage difference variables in Model 2, since in such way we can estimate the effect of the inter-country wage difference at constant effect of the intra-country wage difference. We also estimated the semiparametric response model accounting for endogeneity of the variable on the inter-country wage difference, which we refer to as Model 2-IV. We also report the average partial effects of all model variables on predicted individual job dissatisfaction.

Both Model 1 and Model 2 that were estimated using the parametric ordered probit show the positive and statistically significant correlation between individual age and job satisfaction. Since the variable on age represents an individual experience in our model, people, who have been working a longer time with their current employer, are more satisfied. Consequently, there is a negative average partial effect of age on the probability of a working individual being dissatisfied with her job. Coefficients on the number of children in both parametric models have negative signs and are statistically significant at the 95% level: more children in families requires parents to spend more time with them, while under the fixed working time framework at the majority of employers in Tajikistan, parents cannot easily choose to increase their spare time, which increases their job dissatisfaction. The size of durable goods owned by families increases the individual job satisfaction. Since the current value of durable goods owned by families

represents their wealth status, having more wealth makes people happier. The average partial effect confirms this: being wealthier decreases the probability of job dissatisfaction. Variables which define individual education are positively correlated with job satisfaction: both variables on education from vocational schools and university have positive coefficients, and negative average partial effects on the probability of job dissatisfaction. While the estimated coefficient on the highest level of education from technical schools is not statistically significant, the coefficient on university level of education is statistically significant at 95% level in both models. The affiliation of an employer with a social security scheme has a positive correlation with a worker's job satisfaction: the coefficient on the variable is statistically significant at 95% and 99% significance levels in Model 1 and Model 2, respectively. Since the affiliation with social security scheme guarantees employees both social and health benefits, it decreases the probability of workers' job dissatisfaction. The gender difference does not have a significant effect on individual job satisfaction: coefficients on male gender dummies are not significantly different from zero in both models. Living in the capital city does not have a significant impact on the individual job satisfaction in Model 1, but with inclusion of the inter-country wage difference in Model 2, the statistical significance of its coefficient rises to 90% level. Working in fixed premises does not have a significant impact on individual job satisfaction. Unlike working in fixed premises, working in streets or markets decreases individual job satisfaction at 99% significance level in both model specifications. The average partial effect of this variable on the probability of job dissatisfaction is positive.

Both variables, inter- and intra-country wage differences, have negative correlations with the job satisfaction in parametric Models 1 and 2. The coefficients on the intra-country wage difference are statistically significant at 99% significance level across both model specifications, even despite of inclusion of the variable on the inter-country wage difference. Such result is

consistent with the job satisfaction literature: an individual job satisfaction increases with wage residuals. Since we use the reverse of residuals the sign of its estimates is also reversed. The average partial effect of this difference is positive, which implies that the probability of job dissatisfaction increases if people receive wages lower than they could receive at similar jobs at other employers within Tajikistan. Adding the variable on the inter-country wage difference not only increases McFadden's pseudo-$R^2$ of the model, but also shows a significant correlation of this variable with job satisfaction. The coefficient on the inter-country wage difference is negative and statistically significant at 95% level. Its estimate in the Model 2 shows that, even keeping the effect of individual intra-country wage difference constant, the difference between expected wages from migration to another country and current wages in the home country reduces significantly the job satisfaction of working migrants' relatives in the home country. The average partial effect of this variable is positive implying that an access to the information on outside wage distribution increases the probability of workers' job dissatisfaction.

Table 1.4 reports both estimates and average partial effects of the semiparametric ordered response models. The semiparametric estimation is based on the index representation of the model variables with normalized coefficients by coefficient of one of model's continuous variables. Since parametrically estimated models show a positive and statistically significant estimate on the *age* variable, we normalized other coefficients of the model using this variable.[4] The positive coefficient on age allows us to consistently estimate signs of normalized coefficients of other variables. Its significance allows ratios of other coefficients with respect to it to be finite. Taking into account the positive relationship between individual age and reported job satisfaction, we expect that the signs of the semiparametric estimates of variable coefficients in the Model 1

[4] In order to satisfy the identification condition C.3b in Klein and Spady (1993), we did not include other functions of age.

and the Model 2 would have the same signs as of the parametric coefficient estimates of corresponding variables. Table 1.4 shows that variables in the semiparametric Model 1 and Model 2 have the same sign effects on individual job satisfaction as in parametric models. Interesting to note is that the significance of the coefficient estimates of variables on number of children (*ch14*), the value of durable goods (*durs*), education level from vocational schools (meduc) and universities (heduc), living in the capital city (*capl*) and working in fixed premises (*fdpl*) increases to 99% significance level in semiparametrically estimated Models 1 and 2, which might be due to our relaxed distributional assumptions.

Using estimates of variable coefficients, we estimate the average partial effects of the continuous explanatory variables in the semiparametric models of structural equations as the sample average of differences between the semiparametric expectation of the job dissatisfaction conditioned on the model's index where a variable of interest is increased by one keeping other variables fixed, and the semiparametric expectation of the job dissatisfaction conditioned on the index which is estimated with initial values of the variables. The average partial effects of the dichotomous variables in the semiparametric models of structural equations is also estimated as the sample average of differences of two semiparametric expectations of job dissatisfaction, where the first expectation is calculated conditionally on the index where the variable of interest is set to 1, and where the second expectation is conditioned on the index where the same variable is set to 0, while remaining variables in both indexes are kept fixed. Sizes of the average partial effects of variables on the probability of job dissatisfaction in all three semiparametric models are reported in last three columns in the Table 1.4. Their absolute values differ from those of estimated using the corresponding parametric models, i.e. Model 1 and Model 2. However, the sign effects of average partial effects of variables on the probability of job dissatisfaction are the same across all models.

**Table 1.3. Ordered Probit Model: Estimates and Average Partial Effects for Job Dissatisfaction**

Dependent Variable : Job satisfaction (1-"dissatisfied", 2-"neither satisfied nor dissatisfied", 3-"satisfied")

| Variables | Model Coefficients | | Average Partial Effects of Job Dissatisfaction | |
|---|---|---|---|---|
| | Model 1 | Model 2 | Model 1 | Model 2 |
| age | 0.0095 *** | 0.0100 *** | -0.0015 | -0.0016 |
| | (0.0026) | (0.0026) | | |
| resm | - | -0.1249 ** | - | 0.0202 |
| | | (0.0562) | | |
| resw | -0.4076 *** | -0.3815 *** | 0.0662 | 0.0618 |
| | (0.1160) | (0.1149) | | |
| ch14 | -0.0323 ** | -0.0325 ** | 0.0052 | 0.0053 |
| | (0.0153) | (0.0154) | | |
| durs | 0.0091 ** | 0.0090 ** | -0.0015 | -0.0015 |
| | (0.0044) | (0.0045) | | |
| meduc | 0.0623 | 0.0560 | -0.0098 | -0.0089 |
| | (0.0827) | (0.0828) | | |
| heduc | 0.2199 ** | 0.2137 ** | -0.0319 | -0.0311 |
| | (0.1076) | (0.1071) | | |
| male | 0.0620 | 0.0559 | -0.0101 | -0.0091 |
| | (0.0584) | (0.0586) | | |
| capl | -0.1413 | -0.1588 * | 0.0248 | 0.0281 |
| | (0.0867) | (0.0869) | | |
| ssec | 0.1739 ** | 0.1785 *** | -0.0276 | -0.0282 |
| | (0.0679) | (0.0679) | | |
| fdpl | 0.1170 | 0.1167 | -0.0183 | -0.0182 |
| | (0.0826) | (0.0825) | | |
| smpl | -0.3321 *** | -0.3400 *** | 0.0623 | 0.0639 |
| | (0.0834) | (0.0835) | | |
| Constant-cut1 | -1.0024 *** | -1.0198 *** | - | - |
| | (0.1145) | (0.1147) | | |
| Constant-cut2 | -0.1149 | -0.1307 | - | - |
| | (0.1080) | (0.1082) | | |
| Observations | 3022 | 3022 | 3022 | 3022 |
| Pseudo $R^2$ | 0.030 | 0.032 | - | - |

Standard errors in parentheses, * p<.10, ** p<.05, *** p<.01

**Table 1.4. Semiparametric Ordered Response Model: Estimates and Average Partial Effects for Job Dissatisfaction**

Dependent Variable : Job satisfaction (1-"dissatisfied", 2-"neither satisfied nor dissatisfied", 3-"satisfied")

| Variables | Model Estimates | | | | Average Partial Effects of Job Dissatisfaction | | |
|---|---|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 2 - IV | IV equation: resm | Model 1 | Model 2 | Model 2-IV |
| age | 1 | 1 | 1 | 0.0018 *** | -0.0007 | -0.0009 | -0.0010 |
| | | | | (0.000005) | | | |
| resm | - | -5.5609 ** | -10.2873 *** | - | - | 0.0056 | 0.0135 |
| | | (2.3628) | (2.2861) | | | | |
| resw | -39.0520 *** | -28.0675 *** | -33.4461 *** | 0.0588 *** | 0.0264 | 0.0918 | 0.1108 |
| | (6.3971) | (2.9297) | (8.3216) | (0.00014) | | | |
| ch14 | -3.2695 *** | -2.7482 *** | -2.0265 ** | 0.0016 *** | 0.0022 | 0.0026 | 0.0021 |
| | (1.1662) | (0.5864) | (0.8636) | (0.00004) | | | |
| durs | 1.6101 *** | 1.1490 *** | -0.1732 | 0.0004 *** | -0.0011 | -0.0010 | 0.0002 |
| | (0.3167) | (0.1866) | (0.2552) | (0.000004) | | | |
| meduc | 27.0760 *** | 11.1513 *** | -0.3995 | 0.0034 *** | -0.0191 | -0.0095 | 0.0004 |
| | (8.9724) | (4.1646) | (3.4744) | (0.00015) | | | |
| heduc | 39.7696 *** | 20.0866 *** | 13.6733 *** | -0.0134 *** | -0.0210 | -0.0157 | -0.0141 |
| | (9.7995) | (4.3825) | (4.8604) | (0.00021) | | | |
| male | 2.3587 | 7.5175 *** | 1.5381 | -0.0268 *** | -0.0016 | -0.0064 | -0.0016 |
| | (2.8763) | (2.4411) | (2.2096) | (0.00012) | | | |
| capl | -36.2083 *** | -6.3719 | -0.3707 | -0.0282 *** | 0.0766 | 0.0066 | 0.0004 |
| | (9.4999) | (4.2258) | (3.2393) | (0.00012) | | | |
| ssec | 8.9270 ** | 4.6641 * | 12.5827 *** | 0.0062 *** | -0.0059 | -0.0037 | -0.0142 |
| | (3.7016) | (2.6711) | (3.3824) | (0.00010) | | | |
| fdpl | 28.3550 *** | 17.5706 *** | 9.7133 *** | -0.0227 *** | -0.0175 | -0.0152 | -0.0109 |
| | (5.6926) | (3.7489) | (3.6368) | (0.00017) | | | |
| smpl | -16.7006 *** | -15.3807 *** | -25.6697 *** | -0.0348 *** | 0.0108 | 0.0161 | 0.0242 |
| | (5.6570) | (3.5284) | (4.0473) | (0.00010) | | | |
| netw | | | | 1 | - | - | - |
| Observations | 3022 | 3022 | 3022 | 3022 | 3022 | 3022 | 3022 |

Standard errors in parentheses, * p<.10, ** p<.05, *** p<.01

In addition to semiparametric Model 1 and Model 2, we estimated the Model 2-IV, where we controlled for endogeneity of the variable on the inter-country wage difference. The endogeneity issue of this variable rises because of the family's selection into migration. It might be also due to the possible simultaneity relationship with the job satisfaction: workers dissatisfied from their current wages might decide to send their relatives abroad in order to compensate in lower work earnings. We did not report the endogeneity correction described by Wooldridge (2010), which is based on the two stage Rivers and Vuong (1988) control function approach, where, at the first stage, the reduced form equation for endogenous variable is estimated, then, at the second stage, residuals from the reduced model should be added into the structural ordered response model in order to control for endogeneity of the variable of interest. This approach is based on the strong distributional assumption that the reduced form error term is normally distributed. We conducted tests for normality for the distribution of the first stage residuals, Shapiro-Wilk and Skewness-Kurtosis tests, both tests rejected the null hypothesis. Instead we decided to implement the endogeneity correction using the semiparametric estimation, where one does not have to make any distributional assumption on the error terms.

After controlling for the endogeneity of inter-country wage differences, the estimates of the coefficients of model variables on durable goods, the level of education from the vocational schools, and gender become not significantly different from zero. The effect of number of children is also reduced, but remains statistically significant at the 95% significance level. Other coefficients remained statistically significant from zero at the 99% significance level. The absolute size of the average partial effect of the intra-country wage difference increases from 0.0918 to 0.1108 after we controlled for endogeneity of migration related variable. Likewise the absolute sizes of average partial effects of individual age, employer's affiliation with the social security, and working in street or market places, increase after controlling for endogeneity.

Conversely, the absolute sizes of the average partial effects of the remaining variables, number of children, current value of durable goods, education levels, gender, living in the capital city, and working in fixed premises, are lessened after we controlled for the endogeneity of inter-country wage differences.

We also report the first stage estimates of the reduced form equation for the inter-country wage difference (*resm*). The coefficients of all variables estimated in this equation are statistically significant at 99% significance level. Furthermore, even though the reduced form equation is estimated using the semiparametric nonlinear model, we performed the test for weak instrumental variables basing on the F-statistics from the first stage ordinary least squares estimation. The null hypothesis on the weak instrument was rejected.[5]

Now we turn to the effect of the inter-country wage difference. The coefficient on this variable in the semiparametric Model 2 is negative and statistically significant at 95% significance level. After controlling for its endogeneity, the size of its coefficient almost doubles, its significance also increases to 99% level. The average partial effect of the inter-country wage difference on the probability of job dissatisfaction is positive and increases from 0.0056 to 0.0135 after we control for its endogeneity. However, the size of its average partial effect even after controlling the endogeneity remains smaller in the absolute size than its average partial effect estimated using the parametric Model 2. This result indicates that even after keeping the effect of intra-country wage differences constant, the difference between the expected wages from

[5] We looked at whether the F statistic from the first stage OLS estimation is larger than 10 (Staiger & Stock, 1997). The reported F statistics is 32.64, which supports the validity of our instrument. Then, by comparing the Cragg-Donald Wald F statistic (235.202) to the Stock-Yogo weak identification test critical values (10% maximal IV size is 16.38), we were able to reject again the null hypothesis on the weak instrument (Stock & Yogo, 2005).

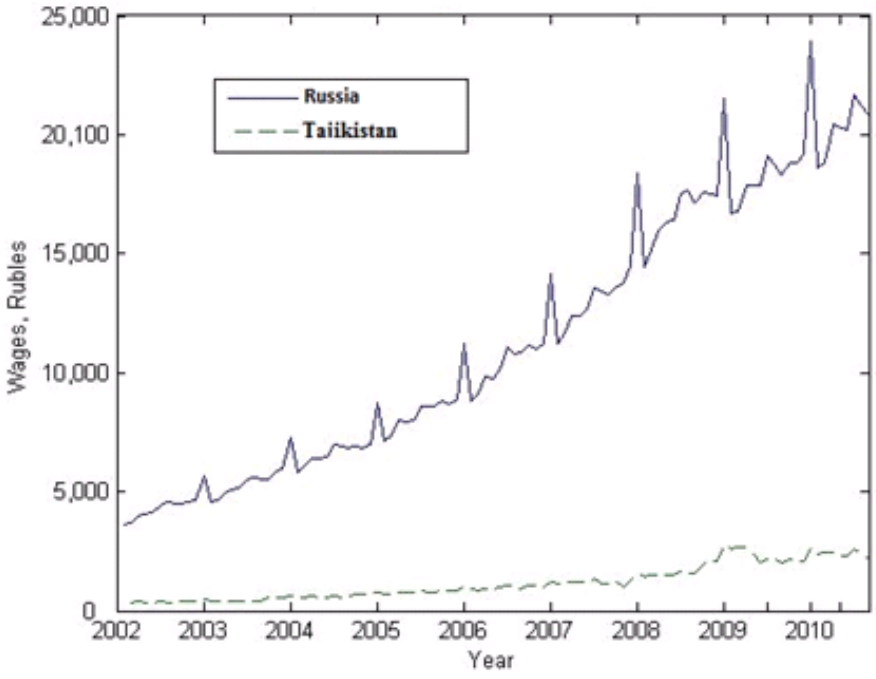
migration and current wages of working members among migrant relatives remaining in the source country increases their dissatisfaction from current jobs.

Such a strongly positive effect of the inter-country wage difference on the probability of job dissatisfaction indicates that it might be destructive for economic development of the source country. Since there is a positive relationship between job dissatisfaction and job quits (for example see Kristensen and Westergaard-Nielsen (2004)), migrant relatives would be more likely to leave their jobs once the gap between the outside wage distribution and the intra-country wage distribution increases. Firms in the source country will be losing workers, consequently, their market competitiveness, due to increasing outmigration. Furthermore, the rigidness in wages in the source country compared to the dynamic wage increase in the destination country will be attracting more migrants to the destination country, living the source country with the shortage of labor. With limited capital endowment, the firms in less-developed countries cannot offer higher wages, hence would be less successful in attracting back migrants.

A similar process is observed in our chosen country case, Tajikistan, with respect to its main migration destination country, Russia. Wages in these countries during Soviet period, when the common market existed, were closer to each other. Schroeder (1981) noted that there were no big differences in average wages of state employees among the Soviet Republics in 1960-1978. Wages in these two former Soviet countries started moving away from each other in early 1980s, and accelerated after the collapse of the Soviet Union. Current developments in restricting the free movement of people such as the introduction of migrant quotas also contributed to speeding up the wage divergence between these countries. Figure 1 shows the scale of the accelerating divergence of Tajik real wages from Russian real wages after the collapse of the Soviet Union. The gap between the average real wages in Russia and Tajikistan increased from 3,335 Rubles in 2002 to 18,600 Rubles by August 2010. In August 2010, the average real wages in Russian

Federation were about 8.5 times higher than those in Tajikistan (Statistical Committee of CIS, 2011; Statistical Agency of Tajikistan, 2011; Russian State Statistical Committee, 2010). Such differences in wages resulted in increasing seasonal labor migration from Tajikistan to Russia, which might be positively related with increasing job quits in Tajikistan.

**Figure 1.1. Comparison of Real Wages between Russia and Tajikistan, Rubles**




*Note:* Data on wages and Consumer Price Index in Tajikistan are from the website of the Statistical Agency of Tajikistan. (2010, December 1). Data on wages and Consumer Price Index in Russia are from Russian State Statistical Committee Monthly Reports on Social and Economic Conditions of Russian Federation (2010). Exchange rates used in converting Tajikistan's wages to Russian Rubbles are from the website of the National Bank of Tajikistan (2011).

### *1.5. Conclusion*

An increasing inflow of remittances is not only destroying the labor participation of remaining members of migrant families, but also increases job dissatisfactions of those who still continue working. Once working migrant relatives in the source country receive information on wage distribution in the destination country through either the size of received remittances or the information received directly from migrants, they are able to build then own expectations on the size of earnings they could receive if they migrated. If the gap between expected wages from migration and current wages increases, working relatives of migrants become dissatisfied with their current jobs.

Using both parametric and semiparametric econometric models, we find a positive significant effect of migration on the increase in the probability of job dissatisfaction of working migrants' relatives in the source country, Tajikistan. The effect remains significant even when we control for possible endogeneity of the migration related variable. Tajikistan has a much lower wage distribution relatively to its main migration destination country, Russia, which attracts more migrants every year from Tajikistan to Russia. An accelerating wage gap between Russia and Tajikistan after the collapse of the Soviet Union not only drives more Tajikistan's population into migration but also increases the job dissatisfaction of those who left behind.

## Part II. Chapter 2. Decisions on Professional Education by Migrant Families

### *2.1. Introduction*

Accumulation of human capital – individual skills, knowledge and capabilities – is vital for economic development. Improvement of the national human capital stock is dependent on individual and household decisions on school attendance. Such decisions depend on factors defining family economic conditions, such as international migration. Our focus in this paper is to study the effect of migration on decisions on acquiring the professional education in the source country. We look at how the migration of one person to another (destination) country impacts on such decisions of his relatives who remain in the source country.

Professional school attendance decisions may be affected differently by the international migration of family members, since international migration of family members changes family habits and resources such as consumption, expenditures and income. From one side, remittances as the main attribute of migration faced by migrant families may have positive effects on education. Remittances encourage families to invest more in education. They help to alleviate financial constraints faced by migrant families, and, consequently, raise investment in education. With increased income, migrant families can afford to pay school fees, transportation and school essentials. In addition, families can hire labor to work in households, family owned businesses and farms, thereby freeing children from doing such work and allowing them to spend more time on education.

On the other side, there are other side effects of migration which negatively impact the decisions of going to schools by non-migrating relatives. The reduction in family consumption because of the migration of economically active family members might decrease schooling if family members need to work more to subsidize the migrant. Families might make this choice if

they see the migration of a family member as an investment which eventually will yield positive returns. Consumption reduction would increase economic hardship in migrant families requiring children to drop out of schools to do housework or market work. In addition to consumption changes, parental absence and expected returns to education give rise to the possibility that migration may also negatively affect decisions on schooling. Parental absence leads to less parental input in children's education. Lower expected returns to education in sending countries compared with low-skilled labor wages in migrant receiving countries might deter school attendance.

We ask how international migration affects the decision of non-migrating family members to acquire education. This is not a new question. Recent work has answered it both positively as the effect of remittances, and negatively as the overall effect of migration which includes its negative side effects. Cox and Ureta (2003) found that remittances have a significantly larger impact than the other income on the school retention, and this effect is still larger in rural than in urban areas of El Salvador. Their results are consistent with consequent findings by other researchers (Acosta, 2006; Calero, Bedi, & Sparrow, 2009). Amuedo-Dorantes and Pozo (2010) found a positive effect of remittances on children's education in the Dominican Republic. Yang (2008) found that exogenous shocks to foreign exchange rates that increased the amount of remittances received by migrant families had favorable effects on the educational attainment of children. The effect is higher for boys with older mother-migrants who migrated for short periods.

Other studies found a negative relationship between migration and educational attainment. McKenzie and Rapoport (2010) using the historical state migration rates as instrumental variables found the negative impact of migration in secondary school age boys and girls in Mexico. They point out that this negative effect of migration is the combination of three

main effects of migration: the positive effect of remittances and two other negative side effects, parent's absence and children's migration prospect . Their result shows that the positive effect of migration has been surpassed by these two negative side effects of migration causing less school attendance of children. Furthermore, Amuedo-Dorantes and Pozo (2010) found that that the positive effect of remittances shrinks if other side effects of migration are allowed into the model.

The estimation methods and model specifications differ in each study. Cox and Ureta (2003) used the Cox proportional hazard model to study the effects of remittance size and receipt on the hazard of school-leaving . Acosta (2006) used both Probit and Instrumental Variable Probit models with binary endogenous explanatory variable on the receipt of remittances by households. Similarly, Calero, Bedi, and Sparrow (2009) used the Instrumental Variable Probit model to study the effect of the size of received remittances on the school enrollment. Amuedo-Dorantes and Pozo (2010) used the Two-Stage Linear Probability model of school attendance by children with an instrumented dichotomous variable for remittance receipts by households . Yang (2008) used a Fixed Effect model to estimate the effect of the change in the size of remittances on changes in fractions of children in the household reported as being students. McKenzie and Rapoport (2010) used different models (Ordinary Least Squares, Two Stage Least Squares, Ordered Probit, Instrumental Variable-Ordered, Censored Probit, and Instrumental Variable-Censored Ordered Probit), where the migration related variable was specified as dichotomous taking the value one if a household has a migrant member and zero otherwise.

These studies do not separate the effects of remittances from migration's other "side effects" . Some authors even argue not to separate the side effects of migration from remittances and treat them as a single variable, i.e. as the product of these two variables in place of them in the structural model:

> "We argue strongly that one actually cannot (in most cases) separate remittances from migration, because these phenomena are intertwined and endogenous. In fact, it is not immediately clear why one would want to separate them and what the pure "impact of remittances" would mean or imply." (McKenzie & Sasin, 2007, p.6)

However, if these other effects matter, ignoring them can result in biased estimates even after correcting for the endogeneity by using instrumental variable approach. This follows as any instrumental variable correlated with remittances or migration will also be correlated with other side effects of migration, which are unobserved and are a part of the error term of the structural model. Therefore, even applying an instrumental variable approach to either remittances or to a migration dummy variable would not solve the endogeneity issue. The appropriate model should include both remittances and side effects of migration in order to correct for the endogeneity of both of these variables.

There are several side effects of migration about which we need to be concerned. Firstly, there is the potential decrease in consumption due to absence of economically active household members because of migration. In order to attain the same level in consumption migrant relatives may choose to work more in the market than they did before the migration. Secondly, the absence of the migrant implies that the remaining family members need to do more work within the household than they did before their relative's migration. Such an increase in work, whether in the market or within household, would require time reallocation from education, therefore some migrant family members might choose to quit schools. Furthermore, the migration of elder members of the household might inspire the younger generation for future migration (a demonstration effect). Children in such families might anticipate migrating after achieving adulthood, and would be less attracted to continuing their education in professional schools. There might be other minor negative side effects of migration or its spillover effects which we do not list here, which jointly with these migration effects can produce a significant effect on individual decisions about education. Such ignorance of the side effects would result in biasness

of estimate on remittances depending whether side effects of migration favor or do not to school attendance.

These side effects of migration are mostly unavailable is existing surveys, and therefore it is difficult to control them in a regression analysis. One way to capture side effects of migration is to include a dummy variable in the model along with remittances. The dummy variable defines whether a household has a current migrant member living and working abroad. Such a variable would explain the change in the intercept between the households with migrants and households without migrants -- with remittances being constant -- which is due to other unobserved side effects of migration.

The inclusion of two variables related to migration in the binary response structural model of the school attendance decision raises the difficulty in econometric estimation. Since both the household's involvement in migration and the size of remittances sent to the household are not random events, our model would contain two endogenous explanatory variables. The complication in estimation rises because our binary response structural model has two endogenous explanatory variables: one is continuous -- the size of remittances received, and another is dichotomous -- a change in intercept caused by other side effects of migration corresponded to self-selection into migration. Because of the complicated estimation nature of such a model, many econometricians suggest for treating only one endogeneity issue. For example, addressing the question on two endogenous variables in a model with one continuous and the other binary, Angrist (2010) answered on his website:

> "Models with multiple endogenous variables are indeed hard to identify and the results can be hard to interpret...So we don't usually like to see them – for one thing it's not clear why you're tackling two causal questions at the same time; one is hard enough...So any time someone shows me a problem with more than one endogenous variable, my first question is always: why?" (Angrist D. , 2010).

Wooldridge (2010) discusses each case separately: a model with only continuous endogenous explanatory variable and a model with only dichotomous endogenous explanatory variable. He suggests three approaches to deal with the case of model with only one continuous endogenous variable (i.e. only remittances): Rivers and Vuong (1988) two stage least squares approach, control function approach or conditional maximum likelihood estimation. Wooldridge suggests to use Bivariate Probit estimation in a case with only binary endogenous variable, i.e. a migration dummy in our model.

We used the Bayesian Limited Information method discussed in Tsurumi (1990) with the data augmentation and Monte Carlo Markov Chain algorithms. Such estimation allows us to specify the structural model of interest, i.e. the decision on professional schooling, along with two reduced form equations for endogenous explanatory variables of the structural equation, i.e. the dummy variable on migration, and a continuous variable on remittances. In our structural model the side effects of migration are defined by our endogenous dummy variable on whether an individual's household has a migrant who left before the start of the agricultural season in 2007. This dummy variable on migration relates to the household's self-selection into migration. Therefore, in its reduced form equation, we use community level migrant networks to control household's decision on migration. To control for endogeneity of the size of remittances sent to the household, we use in its reduced form equation the migrant network and the size of last 12 month harvest. The variable on the current harvest has a strong correlation with the size of remittances, since they might be used to buy fertilizers, seeds, or to expand the size of farming land for the current agricultural season by renting additional spots from neighbors. In contrast, the variable on the current harvest does not have a significant correlation with the decision on migration, since such decisions were made before the start of the agricultural season.

The advantage of using the Bayesian Limited Information Estimation is that it allows us to divide the joint posterior of into two parts: conditional posterior for the structural model and the marginal posterior for the reduced form equations. Such posterior blocking is useful in drawing parameters of each equation separately from their target distributions. Another advantage of this estimation is that using the data augmentation algorithm we can construct the latent variables for the decision on schooling and, importantly, for dummy variable on migration. The constructed variable for the latter would reflects the continuous relationship between the likelihood of household's involvement in migration and possible negative side effects of migration. The monotonically increasing positive number would imply both family's increasing likelihood in sending a migrant and the increase of the costs of acquiring education because of migration. Its decreasing negative number would imply the decrease in the likelihood of family's involvement in migration, and possible benefits of non-migration, as a negative costs in acquiring education because of family's non-involvement in migration.

Results from using such estimation technique in studying the effect of migration on the decisions on schooling might be useful in its policy implications. If there is a positive effect of remittances on schooling, one might be also interested to see the size of the remittances necessary to overcome negative side effects of migration. Such results might be used in developing policies to reduce the negative impact of other negative side effects of migration.

In the next section, we discuss our models of the schooling decision. In the third section we explain the econometric model specification, the Bayesian Limited Information Estimation, and Monte Carlo Markov Chain algorithms used in our regression analysis. Section four discusses the variables and data used in the estimation of the effects of migration on decisions to acquire professional education, as well as its results. We use the data from 2007 World Bank Living Standard Survey on Tajikistan, a small Central Asian, former Soviet and transitional country with

high literacy, which experienced increased labor migration over the last decade. The inflow of remittances reached 35% of country's GDP in 2009 (World Bank, 2011). Such prerequisites make Tajikistan a good country case for studying the impact of migration on educational attainment. The fifth section concludes.

### *2.2. Simple Models on Professional Education*

#### 2.2.1. Model without Migration

Suppose that the parent decides how much to invest in the professional education of his adult child. We use the Becker's human capital model with overlapping generations (Becker, 1991), where each individual lives two periods. In the first period, individual consumption and investment in schooling are dependent on parents' decision. In the second period, the individual receives an income depending on the schooling decision taken in the first period. The parent is interested in his child's earnings in the second period through their altruistic preference, $a$:

$$U(C) + aE_2\,, \tag{2.1}$$

subject to following budget constraint:

$$w_1 \geq C + \gamma s, \tag{2.2}$$

where $U(\cdot)$ is parent's increasing and concave utility function, $w_1$ is the wage rate of parents, $C$ is the composite family consumption in the first period, $E_2$ is child income in the second period; $s$ is a choice of professional schooling, with its price $\gamma$.

Anticipated earnings of the adult children after their graduation from professional schools can be defined as the linear relationship between the wage $w_2$ for low skilled labor, and marginal return, $r$, to the earned human capital from schooling, $\Delta H$:

$$E_2 = w_2 + r\Delta H\,. \tag{2.3}$$

An increment in the human capital is defined by the following production function:

$$\Delta H = f(As)\,, \tag{2.4}$$

as a function of choice of professional schooling, and individual ability. The function $f(\cdot)$ is strictly increasing and strictly concave in its both arguments. Furthermore, without professional schooling attendance the earned human capital from schools would be zero, $f(0) = 0$.

Using equations (2.2)-(2.4), we can rewrite the parent's maximization problem in the following convenient form:

$$\max_{C,s} U(C) + a\big(w_2 + rf(As)\big), \tag{2.5}$$

subject to (2.2). Interior solutions to the problem satisfy following relationship:

$$arAf'(As^*) \le U'(C^*)\,, \tag{2.6}$$

where the strict inequality holds if the left hand side term is equal to zero, for example when $s^* = 0$. According to the equation, the parent chooses to invest in the professional education of their children if the marginal value of the increase of earnings, which depends on children's ability and parental altruism (the left side of the inequality), is equal to the marginal cost of the acquiring professional education (the right side of the inequality) as foregone utility from consumption. Note that if $r$ is very small as in a case of Tajikistan, because of inefficiency of the

professional education system and labor market conditions, then the marginal utility of consumption will be strictly greater than the marginal utility from investing in children education, then the choice of schooling will be zero.

### 2.2.2. Model with Migration

We now consider the decision on whether to obtain professional education or not when one of household's members, not necessarily the parent, migrates to another country in the first period. The household decides on sending its member to another country if the earnings of that member from migration help the household to improve its financial situation. Therefore, the household makes a decision on sending a migrant based on the following choice function:

$$M = 1\{w_B - w_A \geq 0\},$$

where $w_B$ and $w_A$ are household income with migration and without, respectively.

We consider two main attributes of migration which might influence on decision of professional education of non-migrating members of migrant families. The first is the size of remittances received by the household from the migrant, which has a positive effect on schooling. The second attribute of migration is the negative side effect of migration, or monetary and pecuniary costs for acquiring education because of household's involvement with migration (for example see McKenzie and Rapoport (2007)). For example, migrant's absence causes the restructure of household labor supply, and within household workload allocations in such a way that the family might decide not to send their children to universities. Such a negative side effect of migration affects the whole household. The utility function of the parent can be defined as in (2.5). While the budget constraint incorporates both attributes of migration, remittances and costs of migration (i.e. a monetary value of the negative side effects of migration):

$$w_1 + T_m \geq C + \gamma s + R \,,$$

where $T_m$ is remittances, and $R$ is the monetary value of costs in acquiring education faced by the household because of migration. These costs occur only if the household is involved with migration: $R = \{x : \, M = 1, x \geq 0\} = \{x : \, w_B - w_A \geq 0, x \geq 0\}$.

The interior solutions $\{s^*, C^*\}$ to the problem satisfy the same relationship as (2.6) in the model without migration. Using the budget constraint and imposing the condition that parent makes at least a small amount of investment in their children's professional education, $s^* > 0$, we can rewrite (2.6) in the following form:

$$arAf'(As^*) = \gamma U'(w_1 + T_m - \gamma s^* - R) \,.$$

Using this equality we find the following relationships between the decision on professional education with exogenous changes in $\{T_m, R\}$:[6]

$$\frac{ds}{dT_m} = \frac{\gamma U''}{\gamma^2 U'' + arA^2 f''} > 0 \,,$$

and,

$$\frac{ds}{dR} = \frac{-\gamma U''}{\gamma^2 U'' + arA^2 f''} < 0 \,.$$

The first equation shows that the exogenous increase of remittances would imply more schooling for migrant household members. While, the second equation shows a negative effect of costs of migration on schooling: the exogenous increase in migration costs reduces schooling of migrant household members.

[6] The arguments of functions are suppressed for notation simplicity.

## *2.3. Econometric Estimation*

### 2.3.1. Econometric Model Specification

Using our economic model discussed in the previous section, we can define the structural econometric model on the education decision as a binary response model:

$$1(educ_i = 1|\text{data}) = \lambda_0 + \lambda_1 migr_i + \lambda_2 remit_i + X_{1i}\Lambda + \epsilon_i \quad (2.7)$$

where $i$: individual, $educ$ is a binary dependent variable that takes value of 1 if an individual attended school in the last academic year and zero if she did not, $\epsilon \sim N(0, \Omega)$, $X_1$ is a matrix of non-migration related variables affecting the decision to stay in school specified above (including the vector of ones) with a coefficient vector, $\Lambda$. $migr_i$ is a dichotomous variable defining migration costs, and, $remit_i$ is the size of the remittances received by households.

Survey data do not always include variables identifying the negative side effects of migration needed to estimate the costs of migration in acquiring education by migrant household members remained behind. This sometimes makes it impossible to account them in the regression analysis, though ignoring these negative attributes of migration in studying the decision on professional schooling might result in the biasness in estimates of the positive effect of remittances. To overcome this problem we suggest using a dummy variable indicating whether a household has a migrant or not. This way, holding remittances constant, the coefficient on the dummy variable captures other negative side effects of migration, i.e. costs of migration in acquiring education, as a change in the intercept between households with migrants and households without migrants. According to our economic model discussed in the previous section of this paper, households face the costs of migration in acquiring education only if one of their

members migrates to another country. Hence, these negative side effects of migration can be defined as a function of household's decision on migration.

We write the reduced form equations for variables $migr$ and $remit$ as:

$$1(migr_i|Z_i, X_{1i}) = (Z_i, X_{1i})\begin{pmatrix}\Pi_{12}\\ \Pi_{22}\end{pmatrix} + V_{2i}\,, \tag{2.8}$$

$$remit_i = (Z_i, X_{1i})\begin{pmatrix}\Pi_{13}\\ \Pi_{23}\end{pmatrix} + V_{3i}\,,$$

where $Z_i$ is a matrix of excluded exogenous variables defining variable $migr_i$ and $remit_i$. $(\epsilon_i, V_{2i}, V_{3i})$ has a zero mean with a trivariate normal distribution.

While equations for both the costs of migration on education and remittances are expressed in reduced forms, note that their structural form equations might include an explanatory variable on the decision to obtain a professional education. The household's migration decision might be driven by the household's investment decision in its adult children's professional education. Parents or elder male adults in the household choose to migrate in order to earn money for investment into education of other family members. Since we are interested in estimating the impact of remittances and the costs of migration on the household's decision on professional education, we would estimate the structural equation of interest (2.7), but equations (2.8) in the reduced forms, in the way described in Tsurumi (1990). We briefly state in the next subsection how we apply Tsurumi's Bayesian Limited Information Estimation method to our problem.

### 2.3.2. Bayesian Limited Information Estimation

We now derive the Bayesian estimators of the structural equation (2.7). Firstly, we start from reduced form linear equations, where the dichotomous endogenous variables are augmented using the Gibbs Sampler algorithm explained in the next subsection (see a stage 1 in the algorithm):[7]

$$Y_I^* = X\Pi_I^* + V_I^* ,$$

where $Y_I^* = (educ^*, migr^*, remit)$, $X = (Z, X_1)$ is a matrix of all exogenous variables, $V_I^* = (V_1^*, V_2^*, V_3^*)$ are error terms, and $\Pi_I^*$ is a matrix of parameters

$$\Pi_I^* = \begin{bmatrix} \Pi_{11}^* & \Pi_{12}^* & \Pi_{13}^* \\ \Pi_{21}^* & \Pi_{22}^* & \Pi_{23}^* \end{bmatrix}.$$

The data augmentation allows to change the binary endogenous variables to continuous variables, which are used in constructing linear forms of corresponding equations. With constructed linear form equations we can apply the Limited Information Bayesian Estimation.

We can retrieve the equation (2.7), where its dependent variable $1(educ_i = 1|\text{data})$ is replaced by augmented values $educ_i^*$, by post-multiplying the above reduced form equation by the matrix

$$\Delta = \begin{bmatrix} 1 & 0 & 0 \\ -\lambda_1^* & 1 & 0 \\ -\lambda_2^* & 0 & 1 \end{bmatrix}$$

to obtain

[7] Hereon, we use superscript * to refer to augmented endogenous variables and their estimated model parameters.

$$W = X\widetilde{\Pi} + \tilde{V}\ ,$$

where $W = (educ^* - \lambda_1^* migr^* - \lambda_2^* remit, migr^*, remit)$, $\tilde{V} = (\epsilon^*, V_2^*, V_3^*)$, and,

$$\widetilde{\Pi} = \begin{pmatrix} \Lambda^* & \Pi_{12}^* & \Pi_{13}^* \\ \Gamma^* & \Pi_{22}^* & \Pi_{23}^* \end{pmatrix}.$$

This equation is identical to augmented equation of (2.7) where

$$\epsilon^* = V_1^* - V_2^*\lambda_1^* - V_3^*\lambda_2^*\ ,$$

$$\Lambda^* = \Pi_{11}^* - \Pi_{12}^*\lambda_1^* - \Pi_{13}^*\lambda_2^*\ ,$$

and,

$$\Gamma^* = \Pi_{21}^* - \Pi_{22}^*\lambda_1^* - \Pi_{23}^*\lambda_2^* = 0\ ,$$

which is the identification condition.

Assuming the first row of $V_I^*$ is distributed as $N(0, \Omega^*)$, where

$$\Sigma = \Delta'\Omega^*\Delta = \begin{bmatrix} \sigma_{11}^* & \delta^{*\prime} \\ \delta^* & \Omega_{22}^* \end{bmatrix},$$

with $\sigma_{11}^* = Var(\epsilon^*)$, $\delta^*$ is a $2 \times 1$ covariance vector between $\epsilon^*$ and $migr^*$, and, $\epsilon^*$ and $remit$; $\Omega_{22}^*$ is a variance-covariance matrix of the error terms of the reduced form equations (2.8) with augmented data. If $\delta^* = 0$ then $\epsilon^*$ is not correlated with $migr^*$ and $remit$, and there is no need for using a simultaneous equation estimator.

Conditioning on the augmented data, and assuming the flat prior distributions for model parameters, we can write the joint posterior as

$$p(\lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*, \Lambda^*, \Pi_2^*, \sigma_{11}^{*2}(1-\rho^{*2}), \Omega_{22}^* | educ^*, migr^*, remit, \Gamma^* = 0, X, V_2^*, V_3^*)$$

$$\propto [\sigma_{11}^2(1-\rho^{*2})]^{-\frac{n}{2}} |\Omega_{22}^*|^{-\frac{n}{2}} \exp\left[ -\frac{1}{2{\sigma_{11}^*}^2(1-\rho^{*2})} Q(educ^* - \lambda_1^* migr^* - \lambda_2^* remit - X_1\Lambda^* - \lambda_3^* V_2^* - \lambda_4^* V_3^*) \right] \exp\left[ -\frac{1}{2} tr\left( Q\big((migr^*, remit) - X\Pi_2^*\big)\Omega_{22}^{*-1} \right) \right],$$

where $\lambda_3^*$ and $\lambda_4^*$ are coefficients of residuals of reduced form equations (2.8) included in the structural equation (2.7) with augmented data such that $\begin{pmatrix}\lambda_3^* \\ \lambda_4^*\end{pmatrix} = {\Omega_{22}^*}^{-1}\delta^*$, $\Pi_2^* = \begin{bmatrix}\Pi_{12}^* & \Pi_{13}^* \\ \Pi_{22}^* & \Pi_{23}^*\end{bmatrix}$ is a matrix of parameters of reduced form equations of $migr^*$ and $remit$, $\rho^{*2}$ is the canonical correlation of $\epsilon^*$ with $educ^*$, and $Q(\alpha)$ is a quadratic form in $\alpha$.

If we evaluate the posterior conditioned on $\Pi_2^* = \widehat{\Pi}_2^* = (X'X)^{-1}X'(migr^*, remit)$ and $\Omega_{22}^* = \widehat{\Omega}_{22}^* = \frac{1}{n} Q\left((migr^*, remit) - X\widehat{\Pi}_2^*\right)$, then it becomes

$$p\big(\lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*, \Lambda^*, \sigma_{11}^{*2}(1-\rho^{*2}) | educ^*, migr^*, remit, \Gamma^* = 0, \widehat{\Pi}_2^*, \widehat{\Omega}_{22}^*, \hat{V}_3^*, \hat{V}_4^*, X_1\big)$$

$$\propto [\sigma_{11}^{*2}(1-\rho^{*2})]^{-\frac{n}{2}} \exp\left[ -\frac{1}{2\sigma_{11}^{*2}(1-\rho^{*2})} Q\big(educ^* - \lambda_1^* migr^* - \lambda_2^* remit - X_1\Lambda^* - \lambda_3^* \hat{V}_2^* - \lambda_4^* \hat{V}_3^*\big) \right],$$

where $(V_2^*, V_3^*) = (migr^*, remit) - X\widehat{\Pi}_2^*$ .

By blocking the joint posterior, we can draw individually conditional posterior distributions of parameters of the structural model (2.7) and the posterior distribution of parameters of reduced form equations (2.8) with the augmented endogenous binary variables.

### 2.3.3. Gibbs Sampler for Data Augmentation

In the previous subsection we show how the joint posterior of the model parameters conditioned on augmented data could be divided into two parts: the conditional posterior of the parameters of the structural equation and the marginal posterior of the reduced form equations. Such posterior blocking allows us to draw each equation parameters individually. In this section we show that the joint distribution of model parameters and augmenting data can be also divided into two parts, which allow us to draw parameters and to construct latent variables of corresponding endogenous binary variables. By denoting $\Upsilon = \{\lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*, \Lambda^*, \sigma_{11}^{*2}(1-\rho^{*2})\}$ and $\Psi = \{\Pi_2^*, \Omega_{22}^*\}$, the join posterior for model parameters and augmented data can be written as:

$$p(\Upsilon, \Psi, educ^*, migr^* | educ, migr, X, remit, V_2^*, V_3^*)$$
$$= p(\Upsilon, educ^* | educ, migr, \Psi, migr^*, X_1, remit, V_2^*, V_3^*) p(\Psi, migr^* | migr, X)$$

Taking into account that priors $p(\Upsilon)$ and $p(\Psi)$ are constant, consider each term on the right hand side of the equation separately:

$$p(\Upsilon, educ^* | educ, migr, \Psi, migr^*, X_1, remit, V_2^*, V_3^*)$$
$$\propto p(educ | educ^*, \Upsilon, migr, \Psi, migr^*, X_1, remit, V_2^*, V_3^*)$$
$$\times p(educ^* | \Upsilon, migr, \Psi, migr^*, X_1, remit, V_2^*, V_3^*)$$

where

$$p(educ | educ^*, \Upsilon, migr, \Psi, migr^*, X_1, remit, V_2^*, V_3^*)$$
$$= \prod_{i=1}^{n} [1(educ_i^* > 0) educ_i + 1(educ_i^* \leq 0)(1 - educ_i)],$$

i.e. when $educ_i^* > 0$ then $educ_i$ must be one, otherwise zero; and,

$$p(educ^*|\Upsilon, migr, \Psi, migr^*, X_1, remit, V_2^*, V_3^*)$$

$$= [\sigma_{11}^{*2}(1-\rho^{*2})]^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma_{11}^{*2}(1-\rho^{*2})} Q(educ^* - \lambda_1^* migr^* - \lambda_2^* remit - X_1\Lambda^* - \lambda_3^* V_2^* - \lambda_4^* V_3^*)\right].$$

Therefore, the parameters of the structural model of (2.7) can be drawn from the normal distribution. Let $\mu^* = (\lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*, \Lambda^*)'$, and $D = \left(migr^*, remit, X_1, \hat{V}_2^*, \hat{V}_3^*\right)$:

$$\mu^* \sim N(\,(D'D)^{-1}D'educ^*, (D'D)^{-1})$$

Similarly,

$$p(\Psi, migr^*|migr, X) \propto p(migr|migr^*, \Psi, X)p(migr^*|\Psi, migr, X),$$

where

$$p(migr|migr^*, \Psi, X) = \prod_{i=1}^{n}[1(migr_i^* > 0)migr_i + 1(migr_i^* \leq 0)(1 - migr_i)],$$

i.e., the sign of $migr_i^*$ predicts perfectly the value of $migr_i$; and,

$$p(migr^*|\Psi, migr, X) = (\sigma_{22.11}^*)^{-\frac{n}{2}} \exp\left\{-\frac{Q\left(migr^* - X\begin{pmatrix}\hat{\Pi}_{12}^* \\ \hat{\Pi}_{22}^*\end{pmatrix}\right)}{2\sigma_{22.11}^*}\right\},$$

where $\sigma_{22.11}^*$ is the first element in $\Omega_{22}^*$. Hence, the parameter vector of the reduced form equation for $migr$ can be drawn using:

$$\begin{pmatrix}\widehat{\Pi}_{12}^{*}\\ \widehat{\Pi}_{22}^{*}\end{pmatrix}\sim N_{km_1}[(X'X)^{-1}X'migr^{*},(X'X)^{-1}\,].$$

The augmented data for both endogenous dichotomous variables, $educ$ and $migr$ can be drawn independently from the truncated normal distribution centered at fitted variables and unit variance using the Gibbs Sampler procedure: using values for parameter matrix $\Pi_2^*$ , we generate missing observations $migr_1^*,\dots,migr_n^*$ by

$$migr_i{}^{*} = X\begin{pmatrix}\widehat{\Pi}_{12}^{(0)}\\ \widehat{\Pi}_{22}^{(0)}\end{pmatrix}+\varepsilon_{1i}\,,$$

where $\varepsilon_{1i}$ is drawn using

$$\varepsilon_{1i}=\begin{cases}\Phi^{-1}(\mathrm{u}*(1\,-\,\mathrm{p}_1)\,+\,\mathrm{p}_1) & \text{if } migr_i=1\\ \Phi^{-1}(\mathrm{u}*\mathrm{p}_1) & \text{if } migr_i=0\end{cases},$$

and, $u$ is distributed by uniform $U(0,1)$ , $p_1=\Phi\left(-X\begin{pmatrix}\widehat{\Pi}_{12}^{0}\\ \widehat{\Pi}_{22}^{0}\end{pmatrix}\right)$, where $\Phi$ is a standard normal cumulative distribution function.

Given augmented variables $migr^*$, and constructed residuals from the reduced form equations (2.8), $\hat{V}_2^*$ and $\hat{V}_3^*$, as well as the values of parameters $(\lambda_1^*,\lambda_2^*,\lambda_3^*,\lambda_4^*,\Lambda^*)$, we generate the latent variable $educ_1^*,\dots,educ_n^*$ by:

$$educ_i^{*}=\lambda_1^{*}migr_i^{*}+\lambda_2^{*}remit_i+X_{1i}\Lambda^{*}+\lambda_3^{*}\hat{V}_{2i}^{*}+\lambda_4^{*}\hat{V}_{3i}^{*}+\varepsilon_{2i}\,,$$

where

$$\varepsilon_{2i}=\begin{cases}\Phi^{-1}(\mathrm{u}*(1\,-\,\mathrm{p}_2)\,+\,\mathrm{p}_2) & \text{if } educ_i=1\\ \Phi^{-1}(\mathrm{u}*\mathrm{p}_2) & \text{if } educ_i=0\end{cases},$$

and, $p_2 = \Phi\left(-\lambda_1^* migr_i^* - \lambda_2^* remit_i - X_{1i}\Lambda^* - \lambda_3^* \hat{V}_{2i}^* - \lambda_4^* \hat{V}_{3i}^*\right)$.

As it was discussed above, the advantage of using the Gibbs Sampler data augmentation algorithm is that it allows us to change the non-linear form equations into corresponding linear forms, which are estimated using Limited Information Bayesian model. The major advantage, however, is that the constructed latent variable for the dummy variable on migration helps us to estimate the continuous relationship between the likelihood of household's involvement in migration and possible negative side effects of migration. The monotonically increasing positive number would imply both family's increasing likelihood in sending a migrant and the increase of the costs of acquiring education because of migration. Its decreasing negative number would imply the decrease in the likelihood of family's involvement in migration, and possible benefits of non-migration, as a negative costs in acquiring education because of family's non-involvement in migration.

### 2.3.4. Metropolis-Hasting Algorithm with Augmented Data and Random Walk

We apply the Bayesian Limited Information Estimation based on several stages of Monte Carlo Markov Chain (MCMC) algorithms, both the Gibbs Sampler and the Metropolis-Hasting Algorithm with a random walk, to estimate our model on decisions on professional education of migrant family members. We assume that corresponding flat priors for all parameters.

- **Step 1.** Using Gibbs Sampler algorithm derive the augmented data $migr^{*(0)}$, substitute $1(migr_i|Z_i, X_{1i})$ in (2.8). Estimate the system of reduced form equations (2.8) by ordinary least square (OLS) to derive $\hat{\Pi}_2^{*(0)}$ and construct their residuals, $\hat{V}_2^{*(0)}$ and $\hat{V}_3^{*(0)}$. Derive

$$\Omega_{22}^{*(0)} = \left((migr^{*(0)}, remit) - X\hat{\Pi}_2^{*(0)}\right)' \left((migr^{*(0)}, remit) - X\hat{\Pi}_2^{*(0)}\right)/n\,.$$

Then using the Gibbs Sampler algorithm derive $educ^{*(0)}$, substitute $1(educ_i = 1|\text{data})$ in (2.7) and estimate the equation by OLS. Let $\mu^* = (\lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*, \Lambda^*)'$, and $D = \left(migr^*, remit, X_1, \hat{V}_2^*, \hat{V}_3^*\right)$. Calculate

$$\sigma_{11}^{*2(0)}\left(1-\rho^{2(0)}\right) = \left(educ^{*(0)} - D^{(0)}\mu^{*(0)}\right)'\left(educ^{*(0)} - D^{(0)}\mu^{*(0)}\right)/n\,.$$

Construct variance-covariance matrixes for parameters of both structural equation and the reduced form equations: $\Sigma_\mu^0 = \sigma_{11}^{*2(0)}\left(1-\rho^{2(0)}\right)\left(D^{(0)\prime}D^{(0)}\right)^{-1}$ and $\Sigma_\Pi^0 = \Omega_{22}^{*(0)} \otimes (X'X)^{-1}$.

The vector of parameters estimated in this stage would serve as the initial values in estimation by MCMC.

- **Stage 2.** Draw

$$\mu^{*(i)} \sim N\left(\hat{\mu}, s_\mu \Sigma_\mu^0\right),$$

where $\hat{\mu} = \left(D^{(i-1)\prime}D^{(i-1)}\right)^{-1} D^{(i-1)\prime} educ^{*(i-1)}$ , and the scalar $s_\mu$ is set such that the acceptance rate for $\mu^*$ is about 50% . We accept $\mu^{*(i)}$ with probability:

$$\min\left\{\frac{p\left(\mu^{*(i)}, \Pi_2^{*(i-1)}, \sigma_{11}^{*2(i-1)}\left(1-\rho^{*2(i-1)}\right), \Omega_{22}^{*(i-1)}|data\right)}{p\left(\mu^{*(i-1)}, \Pi_2^{*(i-1)}, \sigma_{11}^{*2(i-1)}(1-\rho^{*2(i-1)}), \Omega_{22}^{*(i-1)}|data\right)}, 1\right\}.$$

- **Stage 3.** Draw from inverted Gamma

$$\sigma_{11}^{*2(i)}\left(1-\rho^{*2(i)}\right) \sim IG\left(R_\sigma^{(i)}, df\right),$$

where $R_{\sigma}^{(i)} = \left(educ^{*(i-1)} - D^{(i-1)}\mu^{*(i)}\right)'\left(educ^{*(i-1)} - D^{(i-1)}\mu^{*(i)}\right)$, and $df$ is the degree of freedom.

- **Stage 4.** Draw

$$\text{Vec}\left(\Pi_2^{*(i)}\right) \sim N_{km_1}\left[\text{Vec}\left(\hat{\Pi}_2^{*}\right), s_{\Pi}\Sigma_{\Pi}^{0}\right],$$

where $\hat{\Pi}_2^{*} = (X'X)^{-1}X'(migr^{*}, remit)$, and $s_{\Pi}$ is a scalar set such that the acceptance rate for $\Pi_2^{*}$ is about 50%. Accept $\Pi_2^{*(i)}$ with the probability:

$$\min\left\{\frac{p\left(\mu^{*(i)}, \Pi_2^{*(i)}, \sigma_{11}^{*2(i)}\left(1-\rho^{*2(i)}\right), \Omega_{22}^{*(i-1)} | data\right)}{p\left(\mu^{*(i)}, \Pi_2^{*(i-1)}, \sigma_{11}^{*2(i)}(1-\rho^{*2(i)}), \Omega_{22}^{*(i-1)} | data\right)}, 1\right\}.$$

- **Stage 5.** Draw from inverted Wishart

$$\Omega_{22}^{*(i)} \sim IW\left(R_{\Omega}^{(i)}, df\right),$$

where $R_{\Omega}^{(i)} = \left(\left(migr^{*(i-1)}, remit\right) - X\Pi_2^{*(i)}\right)'\left(\left(migr^{*(i-1)}, remit\right) - X\Pi_2^{*(i)}\right)$.

- **Stage 6.** Derive the augmented data for $educ^{*(i)}$ and $migr^{*(i)}$ using the $i$-th draws parameters of $\mu^{*}$, $\sigma_{11}^{*2}(1-\rho^{*2})$, $\hat{V}^{*}$, $\Pi_2^{*}$ and $\Omega_{22}^{*}$.
- **Stage 7.** Repeat stages two through six a million one hundred thousand times.

We drop initial one hundred thousand draws to eliminate the effects of the starting values. From the remained million, we kept every tenth draw to estimate the distribution of the model parameters. Such large number of replications would allow us to explore the whole posterior and to insure the full convergence in parameter draws.

To confirm that this Limited Information Bayesian Estimation method performs well, we decided to execute the experiment using the above Monte Carlo Markov Chain algorithms. Our experiment shows that elasticity estimates from structural equation, are very close to their true values, they have both negligible biases and smaller variances (see Appendix I for details).

## *2.4. Empirical Analysis*

### 2.4.1. Data and Variable Definitions

We have chosen as a country case Tajikistan, a transitional country which currently experiences a high level of international labor migration. Remittances and migration are playing an augmenting role in lives of Tajik families. The World Bank reports that the size of remittances sent by Tajik migrants reached one-third of the country's GDP in 2009 ranking Tajikistan as the world's most highly dependent country on remittances (World Bank, 2011). As a part of the response to the recognition of current migration trends and other related social issues in Tajikistan, data were collected in 2007 for the World Bank Living Standard Survey (2007 WB LSS) highlighting migrants and their families. This survey includes questions on migration, education, health, labor market, housing, transfers and social assistance, subjective poverty and food security, as well as data for household's expenditure and income. The survey was conducted in two rounds. The first round of data collection was in September and October, 2007 ( during the Ramadan). The second round was in October and November, 2007. There were 4,860 households surveyed in the first round, and 4,490 households were re-visited in the second round.

Tajikistan, at the same time, is a good example country attaining high literacy rates. It inherited a Soviet system of education which requires all children at age 7 to attend elementary schools, and guarantees their education, until the age of 16, in general basic schools (in total 9 years of schooling). The education in public schools is free and parents bear only expenses on

school uniforms, textbooks and other minor payments related to schooling. With enforced free compulsory education and the existence of primary and secondary schools in almost all population units, there will be no significant effect of migration on children education. According to 2007 World Bank Living Standard Survey, there were only 6% of children in age of 8-15, who did not go to schools, or 8% of girls and 4% of boys. And, there is no significant difference between households with and without migrants: in households with a current migrant, 8% of children have not been in school in comparison to 6% of children in households without a current migrant. Therefore, we focus only on the voluntary education which includes education at vocational schools and universities.

After completing the general education at ages of 16-17, a young adult can choose either to continue professional education in vocational schools and universities. The education in vocational schools is four years for those young adults who completed only basic general education and two years for those who completed general secondary education. Education in universities normally takes from four years for Bachelor's degree, and an additional year for the Master's degree.

A current issue in educational attainment in Tajikistan is that a number of students in secondary technical professional schools, or vocational schools, is falling. Tajik youths prefer either to study for higher professions at universities, or choose to work. Some youth looking for higher earnings migrate to Russia to work, since wages in Russia are eight times higher than in those in Tajikistan (Statistical Committee of CIS, 2011). The ratio of students to the number of secondary professional and technical schools dropped from 517 students per school in 1991 to 321 students per school in 2008 (Statistical Agency of Tajikistan, 2011). Therefore, we choose to study how these two developments are correlated whether increased migration in Tajikistan

impacts decisions of young adults in Tajikistan on continuing their education either in vocational schools or universities.

We look at individual decisions on acquiring professional education. Since Tajikistan has high compulsorily general education enforcement rates for children in ages of 7-16, our sample is restricted to groups of males and females in ages of 17-26, who have not received any diploma from either vocational schools or universities. Students with age in this range constitute 90% of those who attended professional schools in the 2006-2007 academic year. Since the survey question asks on school attendance in last academic year, then people at ages 17 and above should have already completed their general education. At age of 27 and above, Tajiks are not willing to study at vocational schools or universities, because at these ages many of them have families and in order to support their families prefer to work rather than study.

Our sample of individuals who have migrant relatives includes those whose relatives migrated before 2007, and currently live and work abroad. We compare the school attendance of this group of people to those who do not have migrants. Observations with later migration patterns were excluded from our sample for two main reasons. Firstly, decisions on migration which are made after decisions on acquiring the professional education in 2006-2007 academic year would not have any effects on education. Secondly, our instrument for the size of remittances, the current harvest, does not have any effect on decisions on migration. Since we restricted our sample to migrants who migrated before 2007, their migration decisions were no affected by following agricultural season and it's any risk factors. Using the variable on harvest we are able to separate the size of remittances from decisions of migration in opposing to the argument by McKenzie and Sasin (2007) on impossibility of their separation (see the first quote in the introduction). We also excluded observations with returned migrants, since according to

our economic model, the migrant families bear costs of migration in acquiring professional education only when they have current migrants.

The decision on school attendance is defined by the dependent binary variable $educ$ as defined in the equation (2.7) above and takes values depending whether an individual attended vocational school or university in the last 2006-2007 academic year. A value of one corresponds to the school attendance. A value of zero refers to individuals who did not attend an academic institution in 2006-2007 and have the highest level of education from secondary schools or below. Individuals who have higher level education than from secondary schools and did not attend academic schools in the preceding academic year were excluded from our sample.

Control variables for this model include (i) individual characteristics: age, ethnic minority, marital status; and, (ii) household characteristics: whether a household has a migrant, the size of monthly remittances both in cash and in-kind (in thousands Somoni), the size of the household, ratio of children to the household size, per capita rate of food consumption deflated by regional prices (in thousands Somoni), parental education, parental death, and variables defining whether the household is located in the capital and other urban areas (the reference is the rural area).

Excluded explanatory exogenous variables in the reduced form equations (2.8) are the community level migrant network (there are 206 population units in our sample) and the income from the harvest in last 12 month converted to a monthly rate (divided by 12 months) in thousands of Somoni. The network variable is commonly used in migration literature. It predicts the likelihood of the household sending migrants abroad. By other words, in communities where the migration is common and has a large rate, the possibility of migration for remaining people increases. To complete our econometric model with multiple endogenous variables, we need an

excluded exogenous variable which separates the size of remittances from the decisions on migration. We use the variable on the current harvest, which is collected during October 2006-October 2007. The current harvest does not have any effect on decisions on migration that were made prior the start of the agricultural season for the current harvest, i.e. before 2007 while the agricultural season in Tajikistan is March-September 2007.

Table 2.1 provides descriptions of each variable, and Table 2.2 provides summary statistics for the whole sample, and individuals in households with and without migrants, as well as corresponding tests on equality in means and variances of these two subsamples. Table 2.2 shows that individuals in families without current migrants have higher attendance rates than individuals in families with migrants: 12.43% of non-migrant family members in age of 17-26 attended schools during the 2006-2007 academic year compared to 5.43% of those in migrant families. The variable on community level migrant network is different for both groups: families who live in communities with large number of migrants are more likely to provide migrants too. Other significant differences at 95% significance level between families with and without migrants are in food consumption, the ratio of children in household's size, ethnicity, urban areas, and the individual gender. Migrant families have relatively larger monthly per capita consumption, which might be due to remittances. The ratio of children in migrant families are larger in mean than in non migrant families. Such difference is due to a smaller size of migrant's household because of migration of some of their members. Since the migrants are predominantly from Tajik ethnicity, the non-migrant families from ethnic minorities are relatively larger in the mean. Furthermore, Tajik migrants are mainly from Tajikistan's rural areas, which shows in non-migrant families' larger mean of being from urban areas. Finally, since the migration in Tajikistan is dominated by men, the mean of male gender in migrant families is significantly smaller than

**Table 2.1. Professional Education Model Variables**

| Variables | Description |
|---|---|
| educ | Binary variable taking a value of one if an individual attended the school on professional education in last 2006-2007 academic year. |
| migr | Binary variable defining whether a household has any current migrant living abroad who migrated before July 2007, 1 if "yes" and zero otherwise. |
| remit | Size of monthly net remittances in the local currency (in thousands Somoni). |
| hhsize | Number of current members living in the household. |
| pcfood | Household's per capita rate of consumption deflated by regional prices (in thousands Somoni). |
| ch14 | Ratio of number of children in the household to the total number of people living in the household. |
| network | Network variable per population unit. |
| harvest | Harvest in last 12 months transferred to a monthly rate (in thousands of Somoni). |
| ethmin | Binary variable defining if an individual is in ethnic minority group, 1 is yes. |
| movoc | Binary variable taking value of 1 if respondent's mother has the highest degree from vocational or special school. |
| mouni | Binary variable taking value of 1 if respondent's mother has the highest degree from university or higher. |
| favoc | Binary variable taking value of 1 if respondent's father has the highest degree from vocational or special school. |
| fauni | Binary variable taking value of 1 if respondent's father has the highest degree from university or higher. |
| capit | Binary variable indicating whether household is in the capital. |
| othurb | Binary variable indicating whether household is in other than capital urban area. |
| modied | Binary variable indicating whether respondent's mother is died. |
| fadied | Binary variable indicating whether respondent's father is died. |
| marry | Binary variable indicating whether respondent is married. |
| men | Binary variable indicating whether respondent is male. |
| age17 | Binary variable taking a value of 1 if a respondent is 17 years old. |
| age18 | Binary variable taking a value of 1 if a respondent is 18 years old. |
| age19 | Binary variable taking a value of 1 if a respondent is 19 years old. |
| age20 | Binary variable taking a value of 1 if a respondent is 20 years old. |
| age21 | Binary variable taking a value of 1 if a respondent is 21 years old. |
| age22 | Binary variable taking a value of 1 if a respondent is 22 years old. |
| age23 | Binary variable taking a value of 1 if a respondent is 23 years old. |
| age24 | Binary variable taking a value of 1 if a respondent is 24 years old. |
| age25 | Binary variable taking a value of 1 if a respondent is 25 years old. |
| age26 | Binary variable taking a value of 1 if a respondent is 26 years old. |
| $v$ | Residuals of the structural form equations, $V_2^*$ or $V_3^*$. |
| sigmasq | Variance of the structural model, $\sigma_{11}^{*\,2}(1-\rho^{*2})$. |
| omeg22 | Variance-covariance matrix of reduced form equations, $\Omega_{22}^*$. |

that in the non-migrant families. The variable harvest does not significantly differs between households with and without migrants. But it is strongly correlated with the size of remittances sent to households.

**Table 2.2. Summary Statistics of Professional Education Model Variables**

| Variable | All Sample | | In Hhs with Migrant | | In Hhs Without Migrant | | Variance ratio test | T-tests |
|---|---|---|---|---|---|---|---|---|
| | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. | | |
| educ | 0.1130 | 0.3167 | 0.0543 | 0.2267 | 0.1243 | 0.3300 | 0 | 0[u] |
| migr | 0.1606 | 0.3672 | 1 | 0 | 0 | 0 | - | - |
| remit | 0.0051 | 0.0183 | 0.0316 | 0.0354 | 0 | 0 | - | - |
| network | 0.0708 | 0.0496 | 0.1055 | 0.0512 | 0.0641 | 0.0464 | 0.0011 | 0[u] |
| harvest | 0.0413 | 0.1546 | 0.0479 | 0.1826 | 0.0400 | 0.1487 | 0 | 0.3049[u] |
| hhsize | 7.8725 | 3.0095 | 7.9271 | 3.0298 | 7.8621 | 3.0060 | 0.7845 | 0.6149[e] |
| pcfood | 0.1051 | 0.0531 | 0.1091 | 0.0500 | 0.1043 | 0.0537 | 0.0208 | 0.0271[u] |
| ch14 | 0.2228 | 0.1555 | 0.2359 | 0.1598 | 0.2203 | 0.1545 | 0.2619 | 0.0190[e] |
| ethmin | 0.1985 | 0.3989 | 0.1628 | 0.3695 | 0.2053 | 0.4040 | 0.0042 | 0.0086[u] |
| movoc | 0.0227 | 0.1488 | 0.0186 | 0.1352 | 0.0234 | 0.1513 | 0.0003 | 0.4154[u] |
| mouni | 0.0112 | 0.1053 | 0.0093 | 0.0961 | 0.0116 | 0.1070 | 0.0006 | 0.5902[u] |
| favoc | 0.0799 | 0.2712 | 0.0899 | 0.2863 | 0.0780 | 0.2682 | 0.0288 | 0.3288[u] |
| fauni | 0.0605 | 0.2385 | 0.0481 | 0.2141 | 0.0629 | 0.2428 | 0.0001 | 0.1154[u] |
| capit | 0.1315 | 0.3380 | 0.0620 | 0.2414 | 0.1448 | 0.3519 | 0 | 0[u] |
| othurb | 0.1340 | 0.3407 | 0.0698 | 0.2550 | 0.1462 | 0.3534 | 0 | 0[u] |
| modied | 0.0441 | 0.2053 | 0.0450 | 0.2074 | 0.0439 | 0.2049 | 0.6829 | 0.9046[e] |
| fadied | 0.1106 | 0.3136 | 0.1302 | 0.3368 | 0.1068 | 0.3089 | 0.0036 | 0.1013[u] |
| marry | 0.3324 | 0.4711 | 0.3411 | 0.4744 | 0.3308 | 0.4706 | 0.7760 | 0.6102[e] |
| men | 0.4373 | 0.4961 | 0.3752 | 0.4845 | 0.4491 | 0.4975 | 0.3970 | 0.0005[e] |
| age17 | 0.1228 | 0.3282 | 0.1209 | 0.3263 | 0.1231 | 0.3286 | 0.8275 | 0.8773[e] |
| age18 | 0.1041 | 0.3054 | 0.0915 | 0.2885 | 0.1065 | 0.3085 | 0.0309 | 0.2312[u] |
| age19 | 0.1101 | 0.3130 | 0.1070 | 0.3093 | 0.1106 | 0.3137 | 0.6518 | 0.7849[e] |
| age20 | 0.1267 | 0.3327 | 0.1163 | 0.3208 | 0.1287 | 0.3350 | 0.1638 | 0.3834[e] |
| age21 | 0.1140 | 0.3179 | 0.1054 | 0.3073 | 0.1157 | 0.3199 | 0.1965 | 0.4525[e] |
| age22 | 0.1026 | 0.3035 | 0.1023 | 0.3033 | 0.1026 | 0.3035 | 0.9918 | 0.9808[e] |
| age23 | 0.0911 | 0.2878 | 0.1023 | 0.3033 | 0.0890 | 0.2848 | 0.0343 | 0.3021[u] |
| age24 | 0.0817 | 0.2739 | 0.0837 | 0.2772 | 0.0813 | 0.2733 | 0.6325 | 0.8359[e] |
| age25 | 0.0764 | 0.2657 | 0.0946 | 0.2929 | 0.0730 | 0.2601 | 0.0001 | 0.0812[u] |
| age26 | 0.0705 | 0.2560 | 0.0760 | 0.2652 | 0.0694 | 0.2542 | 0.1572 | 0.5514[e] |
| Observations | 4016 | | 645 | | 3371 | | | |

[u] T-test with unequal variances.
[e] T-test with equal variances.

There are also no significant differences between individuals in families with and without migrants in the size of household, the parental education, the parental death, the marital status, and age at 95% significance level.

Descriptive statistics indicate that there is a significant difference in school attendance between people in migrant and non-migrant families. However, the major problem with descriptive analyses is that they do not imply the causal relationship between migration and education, and may have a smaller predictive power. We now turn to our regression analysis to estimate causal effects of migration on decisions on acquiring the professional education.

### 2.4.2. Regression Analysis

We estimate effects of migration on the decision to acquire professional education. In our model specification, we control for both the costs of migration on acquiring such education by including a dummy variable on migrant households and the size of remittances. This specification helps us separate the effects of remittances from the negative side effects of migration on acquiring professional education. According to our economic model, the negative side effects of migration and remittances affect the school attendance in different directions. The positive remittances should encourage families to invest in education, while the negative side effects of migration should in contrast discourage it. Such specification allows us to capture the pure effect of remittances received by families: whether larger remittances would increase the likelihood of continuing education in vocational schools or universities by family members in ages of 17-26. It allows us to study separately the effect of remittances and migration's negative side effects. Since according to our economic model both variables work in opposite directions, we are able to estimate the size of remittances that help to overcome the negative side effects of migration.

Additionally, such separation of remittances from the side effects of migration allow us to avoid any bias problems on their true effect.

Firstly, we estimate the structural model (2.7) with the Probit model using Bayesian methods. The Probit estimates as well as the marginal effects of each variable are reported in Table 2.3. As it was predicted by our economic model, the migration side effects, or costs of migration in acquiring education, have a significantly negative correlation with the decision on school attendance at the 99% level. The sign effect of remittances on decisions on professional education corresponds to our economic model discussion -- it is positive, but not statistically different from zero at the 90% level. To control the possible endogeneity issue of migration related variables in this structural model, we applied the Bayesian Limited Information Estimation basing on Monte Carlo Markov Chain algorithms to structural equation of interest and reduced form equations of migration related variables. Parameter estimates of the model for the structural and reduced form equations, their standard errors and estimates of the parameters' auto regression models of order 1 are provided in Table 2.4.

The estimated coefficient on the size of remittances in the structural equation, after controlling its endogeneity issue, is -12.9204 but not statistically different from zero at the 90% level. Such result implies that there is no significant causal effect of remittances on the decision to obtain professional education. Furthermore, the estimate of the negative side effects is -0.0342, but also not statistically different from zero. Such insignificant results imply that decisions on professional education are driven by other factors than the size of remittances and negative side effects of migration. Some of these other factors are not included in our regression. There might be omitted variables that negatively correlated with decisions on migration and the school attendance, their absence in our regressions increases the significance of migration dummy as we observed in Probit estimation results. After controlling for endogeneity, such negative effect is

eliminated. Another explanation is that all other side effects of migration are gathered at the dummy variable in the Probit model, thereby increasing its value, and significance, as their joint effect on migration. Once we allow for the monotonic relationship between this variable and the decision on migration in our Limited Information Bayesian Estimation, such significance of the variable is now reduced.

**Table 2.3. Binary Reponse Model Regression Results: Structural Equation**

| Variables | Estimates | | | Marginal Effects | | |
|---|---|---|---|---|---|---|
| | Mean | St.Dev. | AR(1) | Mean | St.Dev. | AR(1) |
| migr | -0.3506 | 0.1333 | 0.7288 | -0.0063 | 0.0024 | 0.7244 |
| remit | 0.1685 | 2.6021 | 0.7267 | 0.0244 | 0.3773 | 0.7263 |
| hhsize | -0.0402 | 0.0145 | 0.7134 | -0.0058 | 0.0021 | 0.7070 |
| pcfood | 2.0796 | 0.5447 | 0.5468 | 0.3010 | 0.0787 | 0.5439 |
| ch14 | -0.0650 | 0.2343 | 0.6341 | -0.0094 | 0.0339 | 0.6337 |
| ethmin | -0.0152 | 0.0800 | 0.6423 | -0.0003 | 0.0021 | 0.6419 |
| movoc | 0.7642 | 0.1777 | 0.5194 | 0.0036 | 0.0009 | 0.4998 |
| mouni | 0.6076 | 0.2364 | 0.4843 | 0.0016 | 0.0007 | 0.4806 |
| favoc | 0.1959 | 0.1214 | 0.6431 | 0.0022 | 0.0014 | 0.6411 |
| fauni | 0.5955 | 0.1239 | 0.5651 | 0.0065 | 0.0015 | 0.5506 |
| capit | 0.6549 | 0.0849 | 0.5575 | 0.0182 | 0.0026 | 0.5345 |
| othurb | 0.3647 | 0.0856 | 0.5970 | 0.0079 | 0.0020 | 0.5817 |
| modied | -0.1253 | 0.1766 | 0.7220 | -0.0005 | 0.0008 | 0.7168 |
| fadied | -0.3273 | 0.1110 | 0.6868 | -0.0050 | 0.0016 | 0.6790 |
| marry | -0.8823 | 0.0929 | 0.7254 | -0.0323 | 0.0034 | 0.6348 |
| age18 | 0.9026 | 0.1815 | 0.8033 | 0.0079 | 0.0015 | 0.6675 |
| age19 | 1.4198 | 0.1710 | 0.7913 | 0.0185 | 0.0020 | 0.5485 |
| age20 | 1.5865 | 0.1698 | 0.7986 | 0.0242 | 0.0022 | 0.5187 |
| age21 | 1.6202 | 0.1731 | 0.7953 | 0.0205 | 0.0020 | 0.5371 |
| age22 | 1.2918 | 0.1829 | 0.7939 | 0.0105 | 0.0015 | 0.6194 |
| age23 | 1.3230 | 0.1906 | 0.7964 | 0.0080 | 0.0013 | 0.6460 |
| age24 | 1.0142 | 0.2010 | 0.7913 | 0.0041 | 0.0010 | 0.7113 |
| age25 | 0.8892 | 0.2171 | 0.8088 | 0.0025 | 0.0008 | 0.7688 |
| age26 | 0.8902 | 0.2317 | 0.8177 | 0.0020 | 0.0007 | 0.7896 |
| men | 0.5433 | 0.0638 | 0.6313 | 0.0378 | 0.0043 | 0.5933 |
| constant | -2.6425 | 0.2092 | 0.7824 | - | - | - |
| sigmasq | 1.0131 | 0.0320 | 0.0904 | - | - | - |
| Observations | 4016 | | | | | |

**Table 2.4. MCMC Regression Results: Effect of the Remittances and Side Effects of Migration on Decisions on Professional Education**

| | Structural Eqn for *educ* | | | Reduced Form Eqn for *migr* | | | Reduced Form Eqn for *remit* | | |
|---|---|---|---|---|---|---|---|---|---|
| **Variables** | **Mean** | **St.Dev.** | **AR(1)** | **Mean** | **St.Dev.** | **AR(1)** | **Mean** | **St.Dev.** | **AR(1)** |
| migr | -0.0342 | 0.0348 | 0.3583 | - | - | - | - | - | - |
| remit | -12.9204 | 8.7464 | 0.7351 | - | - | - | - | - | - |
| hhsize | -0.0476 | 0.0155 | 0.8341 | -0.0036 | 0.0107 | 0.7046 | -0.0006 | 0.0001 | 0.4010 |
| pcfood | 2.2669 | 0.5791 | 0.7715 | 1.3047 | 0.5248 | 0.7049 | 0.0116 | 0.0059 | 0.4058 |
| ch14 | -0.0669 | 0.2438 | 0.8179 | 0.3952 | 0.1979 | 0.7023 | 0.0005 | 0.0022 | 0.4063 |
| ethmin | -0.0186 | 0.0852 | 0.8358 | -0.1442 | 0.0688 | 0.7172 | -0.0004 | 0.0007 | 0.4048 |
| movoc | 0.7995 | 0.1810 | 0.7412 | -0.0139 | 0.1861 | 0.7123 | 0.0015 | 0.0020 | 0.4175 |
| mouni | 0.5795 | 0.2467 | 0.7313 | -0.2578 | 0.2779 | 0.7507 | -0.0037 | 0.0028 | 0.4189 |
| favoc | 0.1898 | 0.1253 | 0.8218 | 0.0926 | 0.1001 | 0.6958 | 0.000004 | 0.0011 | 0.4290 |
| fauni | 0.5859 | 0.1280 | 0.7820 | -0.0616 | 0.1229 | 0.7221 | -0.0008 | 0.0013 | 0.4039 |
| capit | 0.5963 | 0.0974 | 0.7742 | -0.6841 | 0.1033 | 0.7952 | -0.0048 | 0.0009 | 0.4069 |
| othurb | 0.3375 | 0.0964 | 0.8077 | -0.4373 | 0.0945 | 0.7859 | -0.0020 | 0.0009 | 0.3997 |
| modied | -0.1316 | 0.1883 | 0.8726 | -0.1281 | 0.1342 | 0.7212 | -0.0003 | 0.0014 | 0.4068 |
| fadied | -0.3066 | 0.1155 | 0.8508 | 0.1836 | 0.0820 | 0.6817 | 0.0022 | 0.0009 | 0.3998 |
| marry | -0.8788 | 0.0975 | 0.8715 | -0.1151 | 0.0682 | 0.7043 | 0.0002 | 0.0007 | 0.4064 |
| age18 | 0.8942 | 0.1854 | 0.9052 | -0.0398 | 0.1107 | 0.7158 | -0.0011 | 0.0012 | 0.4034 |
| age19 | 1.4112 | 0.1766 | 0.8977 | 0.0925 | 0.1103 | 0.7117 | -0.000018 | 0.0012 | 0.4088 |
| age20 | 1.5789 | 0.1776 | 0.9077 | 0.0198 | 0.1100 | 0.7286 | -0.0006 | 0.0011 | 0.4086 |
| age21 | 1.6193 | 0.1781 | 0.9009 | 0.0278 | 0.1134 | 0.7289 | -0.00005 | 0.0012 | 0.4164 |
| age22 | 1.2935 | 0.1868 | 0.8998 | 0.0573 | 0.1161 | 0.7213 | 0.0001 | 0.0012 | 0.4057 |
| age23 | 1.3176 | 0.1947 | 0.8974 | 0.1804 | 0.1178 | 0.7062 | 0.0001 | 0.0013 | 0.4076 |
| age24 | 0.9962 | 0.2066 | 0.9031 | 0.0771 | 0.1255 | 0.7173 | -0.0008 | 0.0014 | 0.4072 |
| age25 | 0.8887 | 0.2233 | 0.9103 | 0.2313 | 0.1254 | 0.7002 | 0.0027 | 0.0014 | 0.4161 |
| age26 | 0.8715 | 0.2410 | 0.9168 | 0.1665 | 0.1357 | 0.7162 | -0.0005 | 0.0015 | 0.4151 |
| men | 0.5312 | 0.0662 | 0.8138 | -0.1995 | 0.0561 | 0.7177 | -0.0017 | 0.0006 | 0.4062 |
| $v^{migr}$ | -0.0324 | 0.0353 | 0.3551 | - | - | - | - | - | - |
| $v^{remit}$ | 10.4694 | 8.6739 | 0.7281 | - | - | - | - | - | - |
| network | - | - | - | 9.2802 | 0.5332 | 0.7158 | 0.0748 | 0.0059 | 0.4089 |
| harvest | - | - | - | 0.0199 | 0.1585 | 0.6635 | 0.0094 | 0.0019 | 0.4126 |
| constant | -2.6152 | 0.2456 | 0.8359 | -1.7804 | 0.1439 | 0.7220 | 0.0046 | 0.0015 | 0.4096 |
| sigmasq | 1.0155 | 0.0324 | 0.0978 | - | - | - | - | - | - |
| $omeg22^{migr}$ | - | - | - | 1.0141 | 0.0277 | 0.1350 | - | - | - |
| $omeg22^{migr,remit}$ | - | - | - | 0.0067 | 0.0003 | 0.0185 | 0.0067 | 0.0003 | 0.0185 |
| $omeg22^{remit}$ | - | - | - | - | - | - | 0.0003 | 0.000006 | 0.0039 |
| Observations | 4016 | | | 4016 | | | 4016 | | |

**Table 2.5. MCMC Regression Results: Marginal Effects of Variables in the Structural Equation**

| Variables | Mean | St.Dev. | AR(1) |
|---|---|---|---|
| migr | -0.0049 | 0.0050 | 0.3576 |
| remit | -1.8646 | 1.2616 | 0.7330 |
| hhsize | -0.0069 | 0.0022 | 0.8294 |
| pcfood | 0.3272 | 0.0831 | 0.7677 |
| ch14 | -0.0096 | 0.0352 | 0.8165 |
| ethmin | -0.0004 | 0.0022 | 0.8342 |
| movoc | 0.0037 | 0.0010 | 0.7282 |
| mouni | 0.0015 | 0.0007 | 0.7276 |
| favoc | 0.0021 | 0.0015 | 0.8206 |
| fauni | 0.0064 | 0.0015 | 0.7721 |
| capit | 0.0168 | 0.0028 | 0.7614 |
| othurb | 0.0074 | 0.0022 | 0.8019 |
| modied | -0.0005 | 0.0008 | 0.8680 |
| fadied | -0.0046 | 0.0016 | 0.8466 |
| marry | -0.0318 | 0.0036 | 0.8195 |
| age18 | 0.0078 | 0.0016 | 0.8287 |
| age19 | 0.0183 | 0.0021 | 0.7523 |
| age20 | 0.0241 | 0.0023 | 0.7504 |
| age21 | 0.0205 | 0.0020 | 0.7463 |
| age22 | 0.0105 | 0.0016 | 0.8099 |
| age23 | 0.0079 | 0.0014 | 0.8130 |
| age24 | 0.0040 | 0.0010 | 0.8563 |
| age25 | 0.0024 | 0.0008 | 0.8906 |
| age26 | 0.0020 | 0.0007 | 0.8999 |
| men | 0.0371 | 0.0045 | 0.7920 |
| $v^{migr}$ | -0.0047 | 0.0051 | 0.3539 |
| $v^{remit}$ | 1.5111 | 1.2525 | 0.7266 |
| Observations | | 4016 | |

Household size has a significantly negative correlation with the decision to attend school. Large families face financial constraints; therefore they choose to spend their limited resources on family consumption but not on schooling. Individuals from families who are able to spend more on food consumption are more likely to attend schools. Greater consumption spending might imply that families have sufficient funds and resources to pay for education. Mother's education also increases the likelihood of attending the professional schools. Father's education from

universities has also a positive impact on school attendance. Therefore, parents with higher education would choose to invest in the education of their children. Furthermore, living in urban areas, whether in the capital or other urban areas, where the majority of vocational schools and universities are located in Tajikistan, increases one's likelihood of continuing education in such schools. Men are also more likely to continue the education at universities or vocational schools than women. Father's death, marriage and individual age are negatively correlated with the school attendance decision. With parental death, children enter the labor market earlier or help families with the workload within the household, and would not continue education. Marriage is associated with the increase in one's responsibilities and expenses, therefore, married individuals would choose to do work either inside the household or market work, and therefore would not continue their education. The decreasing size coefficients on age dummies would imply that individuals become less attracted to education if become elder. Remaining coefficients in the model on variables of number of children in the household, ethnicity, and whether individual's mother is died, and father's education from vocational school are not statistically significant from zero at 90% significance level.

The first instrumental variable in the reduced form equations, the variable on the community level migrant network, has a positive and statistically significant correlation with the size of remittances and the decision on migration. This means that households in communities with larger outmigration trends are more likely to choose to send their members abroad. As we expected the third instrumental variable on the current harvest does not have a significant correlation with the decision on migration. Since according to our constructed sample, the migration decisions are taken before the start on the agricultural season of the current harvest. By other words, the migration pattern which might be driven by any current agricultural shocks are excluded from our sample. In contrast, the current harvest has a significantly positive correlation

with remittances. This might imply that migrant families spend remittances for buying fertilizers and high quality seeds, acquiring more land for farming and hiring additional labor, etc. Therefore, the receipt of remittances helps migrants' families achieve higher yields.

The final table, Table 2.5, provides the marginal effects of the variables in the structural equations, with their standard deviations, as well as the coefficient of the auto-regression of order 1. Both marginal effects on the decision on acquiring the professional education are negative but not statistically different from zero.

### *2.5. Conclusion*

In studying the effects of migration on the education, one needs to separate different attributes of migration. Some of them might favor increased educational attainment of migrant family members, others might have negative impact, and still others may not have any effect. Survey data does not necessarily provide the desired information on migration attributes. Therefore, in addition to the attribute of particular interest, one can include a dummy variable to capture the negative "side effects" of migration, i.e. costs of migration on acquiring the education, as a change in the intercept between households with migrants and households without migrants. In such a way we can derive the proper inference about our attribute of interest.

Using the Bayesian Limited Information Estimation to overcome multiple endogeneity issue in the model, we separated the negative side effects of migration from the effect of the size of remittances on the decision on acquiring professional education in Tajikistan. We were able to separate the decision of migration from the remittances in our regression analysis by using the variable on the current harvest, as well as restricting our sample to those who migrated prior to the current agricultural season's yields. Our estimation shows that there is not significant effects of migration and its side effects. The coefficients on both variables, remittances and the dummy

variable on migrant families, have the same, negative signs but are not statistically different from zero. We compared our estimates with those from the Probit model, where the endogeneity issue of migration related variables was ignored. The Probit results correspond to our economic theoretical model: the coefficient on remittances is positive but not statistically significant from zero, while the side effects of migration are negatively correlated with the decisions on acquiring professional education. Such difference in results might be due to two reasons. Firstly, the estimates in the Probit model might be based due to omitted variable bias, because we did not control for endogeneity of migration related variables. Secondly, since all other side effects of migration are gathered at the dummy variable in the Probit model, which increases both its value, and significance. Our applied Limited Information Bayesian Estimation allows to construct latent variables of binary endogenous variables, and to estimate the monotonic relationship between these variables, which reduces the significance of the variable on migration decision.

## Part II. Chapter 3. Migration as a Substitute for Informal Activities

### *3.1. Introduction*

All economies contain some form of informal/ unreported activity. This paper considers the influence migration has on this type of activity. We argue that migration and informal sector activity are viable options for the household.

The migration literature going back to at least Harris and Todaro (1970) and the papers by, for example, Fields (1975, 1976, 1979) and Gang and Gangopadhyay (1987a,b,c), generally introduce the informal sector as a complement to migration – that is, the informal sector becomes a staging ground for those trying to get formal sector jobs, part of the process that drives modern economic growth and, frequently, urbanization. In these models informal sector work and migration are complementary: migrants have difficulty finding employment in formal work in "new" places, so many of them end up informally employed. The informal sector is in the migrant's destination location, along with the good jobs the migrant is hoping to get. It is also possible for the informal sector – if it pays enough – to be the migrant's desired employment.

Our approach is somewhat different as we consider informality and migration as possible alternatives to one another. While the informal sector may be part of the process of economic growth and growing urbanization described in the previous paragraph, the informal sector may also be a home for entrepreneurs, a place to supplement "regular" earnings, or, alternatively, a home of last resort where the vulnerable end up during periods of economic hardship. This is local informal activity and this is the focus of our investigation into a trade-off between migration and informality. As substitutes, migration may effectively "crowd out" informality: migrant's earnings help improve families' finances encouraging their members to be less involved in informal employment. This structure has not been generally addressed in migration models.

Thorbecke (1999) describes the co-existence of modern and informal/traditional sectors in both urban and rural regions, modeling their linkages via social accounting matrices. Building on the Harris and Todaro (1970) model, Gang and Gangopadhyay (1987c) allow for regular and informal employment in both urban and rural regions, with the possibility of open urban unemployment. With this extra complexity, whether migration out of the rural region and rural informality are substitutes or complements depends on relative wages and the various labor supply and demand elasticities.

Using the Living Standards Survey (LSS), our stage is Tajikistan, a poor Central Asian economy and former Soviet Republic possessing both a very large informal sector and extensive external migration. Our aim is to define the direction of correlation between informality and migration, and to examine nuances of the relationship. In the next section we provide a short introductory background to the major economic events in Tajikistan's recent history, emphasizing elements that are important to our story. We then discuss our approach to measuring informal sector activity; we discuss the data used in this study, report on the results and draw our conclusions.

### *3.2. Background*

Tajikistan underwent severe economic, social and political changes following its separation from the USSR. Independence in 1991, with its rupture of economic ties, was followed by civil war among rival regional clans from 1992 to 1997 and then an initially tenuous peace. Tajikistan's GDP fell by 65% from US$2.6 billion in 1990 to US$921.8 million in 1997, while inflation peaked at 1207.2% using the GDP deflator in 1993, two years after independence, and was still at 65.2% in 1997 (World Bank, 2011).

After reaching reconciliation in 1997, the joint government initiated strict fiscal and monetary policies, along with the privatization of small and medium state owned enterprises, and price and trade liberalization.[8] For the last decade annual real GDP growth has averaged 8.4%, and the inflation rate was also moderated at average annual rate of 20.5% over the decade 2001-2010 (World Bank, 2011). Despite these positive developments, Tajikistan remains the poorest country among former Soviet countries with 47.2% of its population living below the poverty line in 2009 (United Nations, 2011). Average monthly wages were US$83 in 2010; 8.5 times lower that in Russia (Statistical Committee of CIS, 2011). For agriculture, forestry and fisheries, which provide jobs to 50% of the employed population, monthly wages average US$42 (Statistical Agency of Tajikistan, 2011).

The institutional transformation in Tajikistan was slowed by its civil war. The absence or weaknesses of newly established institutions spurred the increase of the informal sector in Tajikistan. Severe economic conditions during the war and post-war recovery period reduced the number employed in state enterprises. Extremely low wages and economic recession drove many employees of state-owned enterprises and kolkhozes (collective farms) to self-employment and migration. Tajikistan's Statistical Agency reports the official unemployment rate as increasing from 0.4% in 1992 to 2.9% in 1998, though this is generally recognized as an understatement. An informal consensus suggests in 1999 the unemployment rate was above 40%, including hidden unemployment (Noda, 1999). Financial constraints for families increased after the loss of savings collected during the Soviet period due to high inflation. Families in Tajikistan were not able to solely rely on wages as a source of income, as they did during the Soviet time. The average monthly wage in 1998 was 8,287 Tajik Rubles (US$9.9 at the official National Bank rate), far

[8] The presidential election and the first multi-party elections were held in Tajikistan in 1999 and 2000 respectively, after reaching the reconciliation between confronting parties in 1997.

less than the internationally recognized subsistence level of "one dollar per day". In 1996, the real monetary income was 38.9 % of the 1991 level (Robertson, 1999). Such conditions led to the increase of the shadow economy and informal sector in Tajikistan.

In 2006, the size of the shadow economy in Tajikistan reached 60.9% of GDP, tax avoidance amounted to about one-third of GDP, and home production of food was 14.7% of GDP, while income from in-kind wages and barter exchange was 13.1% of GDP. Informal employment is common in Tajikistan, with only 46% of household members who are in the labor force employed in formal sector work in 2006. Moreover 45.4% of respondents received income from informal employment that was 2.7 times higher than the income from formal employment (United Nations Development Programme, 2007).

Tajikistan is a country with significant external migration, such that approximately 37% of the labor force is working outside of the country. Most emigrants go to Russia (95.3% of migrants, 2007 World Bank Living Standard Survey (2007 WB LSS)). Increasing migration led to the increasing inflow of remittances into Tajikistan, which in its turn helped to support positive economic growth. Tajikistan became the most remittance dependent country in the world. In 2009, the total received remittances were counted as 35% of its GDP (World Bank, 2011).

According to 2007 WB LSS, the international labor migration from Tajikistan is dominated by men (93.5%), from rural areas (76.4% of all migrants), and ethnically Tajik (81.4% of all migrants). Only 10.7 % of migrants had obtained post-secondary schooling; 76.2% graduated from secondary schools. The majority of current migrants were unemployed, 66.5%, and only 26.6% of migrants were working before migration; and, the remaining were students, pupils or militants. 6.6% of migrants remitted both in-kind and in cash in last 12 months; 74.2% remitted in cash only, and 1.0% remitted in-kind only.

We can draw out of this that Tajikistan was an economy in crisis during most of the first decade of separation from the Soviet Union. Over the second decade the economy has become stable and growing, yet marked by two potentially problematic features: a very large informal sector and extensive emigration. The remainder of this paper analyses the relationship between these two phenomena and examines their implications for households and the economy.

### *3.3. Measuring Informal Sector/ Unreported Activity*

The purpose of this paper is to document the impact migration has on informal and unreported activity. To do so we follow the approach used in Dimova, Gang and Landon-Lane (2006) which looks at income and expenditure information at the household level to determine the amount of informal/ unreported activity for each household. There are many definitions of informal activity including, but not limited to, activity in organizations that have less than 5 employees, activity in organizations that do not use modern production techniques (sometimes referred to as traditional sector employment), employment in activities that do not have employment protections, and employment in organizations that do not have access to formal capital markets. In this paper we do not make a distinction among these definitions but rather look for evidence that a household is spending considerably more than its total income. This, we believe, is a good indicator of the unreported activity in an economy. A large component of this unreported activity is informal sector activity.

To measure the size of unreported activity we turn to income and expenditure data at the household level. Total income is computed as including total receipts from employment, net transfers from government agencies, remittances from household members living away from home, the market value of assets consumed (e.g. livestock, vegetables etc.), and the market value of labor services rendered for which payment was in kind. Total expenditure for a household

includes total payments for good and services consumed, the market value of goods and services consumed where payment was made in kind, the market value of assets consumed, and the value of savings (or asset accumulation). We measure total reported income and total reported expenditures, with the excess of total expenditures over income regarded as unreported income.

There are many reasons why there would be a discrepancy between households' reported expenditures and income, such as non-reporting of informal sector income, memory recollection problems, or problems assigning market prices to in-kind consumption or income. Our analysis looks at the variation in this discrepancy across different households and in particular we look at the differences between households that contain migrants and those that do not. Our assumption is that the only major difference between these households is that households with migrants receive observed remittance income.

We use the household as the unit of analysis since expenditures are difficult to assign to any one individual. While the source of formal sector income can often be assigned to an individual, in keeping with our idea of the informal sector, informal income invariably cannot. Formal sector employees may have a second informal sector job; an apparently non-working member of the household may in fact be employed in the informal sector; or children may be participating in the informal sector. People may participate in both formal and informal activities.

Our approach is different from much recent work on the informal economy, which has followed a paradigm set out by International Labour Organization (ILO) and World Bank staff. The approach is nicely summarized in Perry et al. (2007), and synthesized with some earlier approaches especially in Box 1.1 (page 27). The idea is that there are two main definitional strands: the earlier “productive” and the more recently fashioned “legalistic”. The productive categorization defines informality, as its label implies, by the production attributes of a firm: for

example, a firm might be defined as informal if it employs less than 5 people and uses mechanical power or less than 20 people if it does not use such power. The legalistic categorization essentially distinguishes people who have social protection from those who do not. This approach has been used in labor economics and occasionally in international trade for decades, especially in distinguishing between covered and uncovered sectors. These two ways use information about the firm in which the individual is working to identify whether an individual is working in the informal sector. The approaches overlap with one another in their identification of who is in the informal sector, for example, workers employed by firms having limited capital and offering no formal labor market protections are counted by both approaches.

Our approach overlaps with these categorizations but the overlap is not defined along the same rows and columns useful for comparing the productive and legalistic categorizations. For example, our measure will capture those working in formal jobs as their first job, and who work in informal jobs as second or third jobs. The other approaches have difficulty with second and third jobs, even when reported, as individuals may report industry characteristics of their first job which would make it look to the researcher that they were engaging in formal activity when in fact the majority of their income was sourced from informal activity. On the other hand, our approach does not capture informal activity by a household reporting income equal to expenditure.[9]

The main advantage of our approach is that it allows the use of the rich trove of survey data to examine informality and the link between informality and other aspects of the economy. By looking at the disparity between reported income and reported expenditures as evidence of

[9] Recall and measurement error may also play a role here. To minimize this one could consider the difference between expenditure and income as indicating informal sector activity only if expenditure is significantly more than income (Dimova et.al. (2006)).

informal sector activity, our approach does not need detailed information about the working environment (whether it be firm characteristics or worker protections) in order it to assign individuals (or households) to informal sector activity.

## *3.4. Empirical Analysis*

### 3.4.1. Data used

This study uses the 2007 World Bank Living Standard Survey on Tajikistan.[10] The survey data is based on a representative probability sampling on: (i) Tajikistan as a whole; (ii) total urban and total rural areas, and (iii) five main administrative regions (oblasts) of the country: Dushanbe (the capital), Regions of Republican Subordination (RRS), Sogd Oblast, Khatlon Oblast, and Gorno-Badakhshan Autonomous Oblast (GBAO). This data provides a good basis for our analysis as it incorporates all relevant information on the flow of resources in and out of the household. The data is collected by interviewing 4860 households in two rounds from September to November 2007. The first round of interviews was conducted in September-October 2007, during the Ramadan period. The second round was conducted in October-November 2007 to gather additional information, and, to re-administer food consumption to take into account its changes because of Ramadan.

This survey asks questions on migration, education, health, labor market, housing, transfer and social assistance, subjective poverty and food security, as well as data for household's expenditure and income. Income variables include both cash and in-kind forms of remittances, scholarships, wages and bonuses, individual transfers, social assistance, pensions, income from selling harvest, farm animals and poultry (or their product) and other income.

[10] http://go.worldbank.org/IPLXWMCNJ0

Expenditures include payments for food, education, transportation, payments for health and medication, mortgage payments, house utilities and rent, assistance provided to other individuals, payments for the land use, purchases related with land cultivation and harvesting, purchases of farm animal and poultry breeding, and their food. All income and expenditure variables are converted to monthly equivalent for each household in our estimations.

Table 3.1 reports the definitions of the variables used in the regression analysis. The dependent variable used is the natural logarithm of the ratio of reported expenditures to reported income, where income includes remittances from members of the household living away from the household. As described above the income and expenditure variables are computed from self-reported income and expenditure data that includes good or services given or received in kind. Households that report a larger expenditure than income clearly have unreported income. Based on earlier work this discrepancy includes informal sector income that is not reported in the income reports but does show up in the expenditure reports (see Dimova et.al. (2006)).

We investigate the relationship between remittance income (a household's income derived from a household member working away from home and sending money to the household) and informal sector income. We aim to see if these two sources of income are substitutes or complements. To do so we look to see if the presence of a migrant in a household or a recently returned migrant – that is, a member of the household has left the home and is potentially remitting income – has an impact on the amount of excess expenditure over income. Our assumption here is that this excess expenditure over income, while due to many factors including measurement error and recall error, is mainly due to the presence of unreported income. If a household substitutes informal sector income for remittance income then this would show up as a decrease in the excess expenditure over income.

We find that households with migrants have a lower ratio of expenditure to income. We interpret this as a substitution of unreported informal sector income for reported remittance income. We argue that as the difference is so large it is hard to believe that this drop in excess expenditure is due to systematic differences in potential reasons for an excess expenditure over income such as recall error and mis-pricing of in-kind consumption. Another possible reason for this is that there are systematic differences between migrant and non-migrant households with respect to their savings. It could be argued that migration causes income to increase to an extent that migrant households save some of their reported income thus lowering the observed excess expenditure over income. Of course this is a possibility but for the reasons outlined below is a extremely unlikely event. Tajikistan is poor with a large proportion of the population living on or below the poverty line. The average increase in expenditure shown in Table 3.2 is approximately 25% between returned migrant households and non-migrant households. If all the reduction in excess expenditure is due to increased savings then that would mean households save on average 10% of their income. This is an implausibly high number for a developing country whose population are living close to or below the poverty line.

A second consideration is that while we do not have information regarding household savings in our sample there is evidence from other studies that suggest that very few Tajik households have bank accounts. The survey asks whether the household has a bank account and in the survey 99% of respondents did not have a bank account. Also, the ILO (2010, p.33) reports "It is interesting to note that whether or not one receives remittances appears to have little impact on the likelihood of having a bank account. Of all households who receive remittances, 98% do not have a bank account, while 99% of households who do not receive remittances do not have an active bank account". Moreover, there are no differences between migrant and non-migrant households in terms of house ownership. It is very hard to argue that the observed reduction in the

excess expenditure over income is caused by migrant households saving some of their new income.

**Table 3.1. Informal Sector Model Variables**

| Variable Name | Description |
|---|---|
| log(expenditure /income) | Difference of log of totally reported expenditure and log of totally reported income; the income and the expenditure are defined at monthly rates from all reported sources. |
| Migrant (abroad) | A dummy variable taking a value of 1 if a household has any current migrant who is currently abroad and 0 otherwise. |
| Migrant (returned) | A dummy variable taking a value of 1 if a household has an external migrant who was abroad for less than 12 months and recently returned and 0 otherwise. |
| Borrow | A dummy variable taking a value of 1 if the household borrowed money and 0 otherwise. |
| Vocational | A dummy variable taking a value of 1 if the highest level of education for the head of household is a vocational qualification and 0 otherwise. |
| University | A dummy variable taking a value of 1 if the highest level of education for the head of household is a University degree and0 otherwise. |
| Single | A dichotomous variable taking a value of 1 if the head of household is single and 0 otherwise. |
| No. of children (<15) | Number of children in the household with ages less than 15. |
| No. of Elderly (>65) | Number of elders in the household with ages greater than 65. |
| Ethnic | A dummy variable taking a value of 1 if a head of household is a member of an ethnic minority group and 0 otherwise. |
| Urban | A dummy variable taking a value of 1 if a household lives in urban area and 0 otherwise. |
| Land | A dummy variable taking a value of 1 if the household has access to land and 0 otherwise. |
| Self-employed | A dummy variable taking a value of 1 if any member of the household owns his/her business or farm and 0 otherwise. |
| Professional | A dummy variable taking a value of 1 if a head of the household is employed in a professional occupation and 0 otherwise. |

We start with a model that has the difference in log expenditures to log incomes as the dependent variable and variables indicating whether there is a current migrant or a recently returned migrant in the household as an independent variable. We also include other household characteristics to check the robustness of our regression results. The additional variables used in the regression include whether the household had taken a loan to capture whether it was credit

constrained (*Borrow*), and household demographic variables such as whether the household is a single household, that is the head of the household is single and has never been married (*Single*), the number of children under the age of 15 present in the household and the number of adults over the age of 65 in the household. We also include information as to whether the head of the household is a member of an ethnic minority as we want to allow forth potential that ethnic minorities are discriminated against in the informal economy as well as the formal economy.

We include education indicator variables to control for the possibility that unreported activity may be a function of one's education. The education variables we use is an indicator variable that takes the value of 1 if the highest level of education for the head of the household is a vocational training school *(Vocational)* or a University education (*University*).

The next set of variables describes the type of work undertaken by the head of household. We use the variable (*Self-employment*) to reflect whether anyone in the household is self-employed and the variable (*Professional*) to denote that the head of the household works in a professional job. A feature of Tajikistan is the fact that professionals such as doctors and lawyers are paid low wages by the state and it is standard for them to augment their income by taking clients "off the books" – low wages in Tajikistan have driven professionals look for additional, informal, earnings. Since professionals have wider social networks, access to information and flexible time it is easier to them to open own businesses or provide services to other institutes or people. Such cases are common in developing poor countries. For example, Kinyanjui (2010) discusses a case of disempowered professionals in Kenya who do additional informal work after their normal working time. It is also common in Tajikistan to see a doctor practicing at home after hours, teachers providing after-school tutorship to students, lawyers practicing their clientele beyond their office times, all for "under-the-table" payments. Out-of-pocket payments are also common in hospitals and clinics (Falkingham, 2004). Professionals holding managerial positions

in state agencies and enterprises in Tajikistan might also receive an "unofficial" income in forms of gifts or bribes. The corruption and bribery in Tajikistan is common and it has impacted every sector and level of state agencies (UNDP and the Center for Strategic Studies under the President of the Republic of Tajikistan, 2010). Non-professionals, on the other hand, have no access to bribes, since they work at lower occupations and have fewer opportunities to be involved into informal sector employment due their time and physically intensive work. Finally we include a set of variables that aim to capture the opportunity set of households to find informal sector work. Such variables are whether the household lives in an urban area (*Urban*), and whether the household has access to land for cultivation (*Land*).

**Table 3.2. Summary Statistics of Informal Sector Model Variables**

| Variable | Full Sample | | Migrant (abroad) | | Migrant (returned) | | No Migrant | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. |
| log(expenditure /income) | 0.97 | 1.36 | 0.75 | 1.24 | 0.64 | 1.48 | 1.05 | 1.37 |
| Total Income | 558.06 | 811.05 | 575.14 | 620.72 | 911.26 | 839.24 | 513.42 | 833.69 |
| Total Expenditure | 1371.11 | 1851.37 | 1305.72 | 1651.79 | 1686.79 | 2213.84 | 1347.85 | 1838.14 |
| Migrant (abroad) | 0.17 | 0.37 | 1.00 | -- | 0.15 | 0.36 | -- | -- |
| Migrant (returned) | 0.10 | 0.30 | 0.09 | 0.29 | 1.00 | -- | -- | -- |
| Borrow | 0.05 | 0.23 | 0.06 | 0.27 | 0.07 | 0.29 | 0.05 | 0.22 |
| Vocational | 0.11 | 0.31 | 0.08 | 0.27 | 0.13 | 0.34 | 0.11 | 0.31 |
| University | 0.19 | 0.39 | 0.13 | 0.33 | 0.13 | 0.34 | 0.21 | 0.41 |
| Single | 0.01 | 0.11 | 0.01 | 0.12 | 0.01 | 0.08 | 0.01 | 0.11 |
| No. of children (<15) | 2.20 | 1.70 | 2.04 | 1.74 | 2.48 | 1.75 | 2.20 | 1.69 |
| No. of Elderly (>65) | 0.30 | 0.58 | 0.30 | 0.58 | 0.22 | 0.53 | 0.31 | 0.58 |
| Ethnic | 0.21 | 0.41 | 0.19 | 0.39 | 0.22 | 0.42 | 0.22 | 0.41 |
| Urban | 0.35 | 0.48 | 0.24 | 0.43 | 0.25 | 0.43 | 0.38 | 0.49 |
| Land | 0.65 | 0.48 | 0.74 | 0.44 | 0.78 | 0.41 | 0.62 | 0.49 |
| Self-employment | 0.51 | 0.50 | 0.47 | 0.50 | 0.47 | 0.50 | 0.52 | 0.50 |
| No. of observations | 4391 | | 733 | | 447 | | 3280 | |

Sample summary statistics (means and standard deviations) for each variable are reported in Table 3.2 and Table 3.3. There are two samples used: the first is the sample of all households

who report their positive total income and the second sample is the first sample restricted to those households who report an occupation. Included in this second sample are all those households who work. We see that the mean log ratio of expenditure to income is 0.97 which equates to a mean ratio of excess expenditure to income greater than 2.5. Thus there appears to be a large amount of unreported income in Tajikistan. We also break the sample into those households with a migrant who is currently away from home, a household which has a recently returned migrant, and those households with no migrant. The sample means show that households without any migrants have the largest excess expenditure over income with the households with a recently returned migrant having the smallest ratio of excess expenditure over income. This suggests that unreported income of whatever source is being replaced with reported remittances from migrants.

We observe the same picture when we look at the incomes and expenditures separately. Households with migrants have a higher mean income and also a higher mean expenditure with households with returned migrants having the largest income and expenditure.[11]

Approximately a quarter of all households have a member who has migrated with 17% of households having a migrant who is currently abroad and 10% having a recently returned migrant. Note that some households have both a recently returned migrant and a currently abroad migrant. Very few households borrowed money in the survey period (approximately 5%) and less than 1% were households with a non-married head.

The education levels of migrant and non-migrant households are somewhat different. For the full sample only 19% of the households have a university educated head whereas for households with migrants only 13% of the households are university educated. Thus it appears

[11]This is consistent with the story that migrants who are currently abroad may not have received their full compensation and so their remittances are less than the migrants who have returned and earned their full salary.

that migrant households have lower education than non-migrant households. When we add in households with vocational training (non-university post-secondary education) we observe that 32% of non-migrant households have some form of post-secondary education whereas between 21% and 26% of migrant households have some form of post-secondary education.

**Table 3.3. Summary Statistics: Sample of Reported Occupation by Heads of Households**

| | All | | Migrant (abroad) | | Migrant (returned) | | No Migrant | |
|---|---|---|---|---|---|---|---|---|
| Variable | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. |
| log(expenditure /income) | 0.85 | 1.25 | 0.66 | 1.14 | 0.35 | 1.25 | 0.93 | 1.25 |
| Total Income | 611.27 | 848.36 | 660.53 | 615.52 | 1052.48 | 859.21 | 553.77 | 865.20 |
| Total Expenditure | 1422.88 | 1876.75 | 1406.10 | 1617.43 | 1693.86 | 2092.51 | 1400.00 | 1896.46 |
| Migrant (abroad) | 0.14 | 0.34 | 1.00 | -- | 0.13 | 0.33 | -- | -- |
| Migrant (returned) | 0.10 | 0.30 | 0.09 | 0.29 | 1.00 | -- | -- | -- |
| Borrow | 0.06 | 0.25 | 0.07 | 0.29 | 0.09 | 0.33 | 0.06 | 0.24 |
| Vocational | 0.13 | 0.33 | 0.10 | 0.30 | 0.14 | 0.35 | 0.13 | 0.34 |
| University | 0.25 | 0.43 | 0.18 | 0.38 | 0.18 | 0.38 | 0.27 | 0.45 |
| Single | 0.01 | 0.10 | 0.02 | 0.12 | 0.01 | 0.10 | 0.01 | 0.10 |
| No. of children (<15) | 2.15 | 1.62 | 1.97 | 1.67 | 2.38 | 1.63 | 2.16 | 1.61 |
| No. of Elderly (>65) | 0.12 | 0.37 | 0.14 | 0.38 | 0.13 | 0.42 | 0.12 | 0.36 |
| Ethnic | 0.21 | 0.41 | 0.17 | 0.38 | 0.22 | 0.41 | 0.22 | 0.41 |
| Urban | 0.36 | 0.48 | 0.22 | 0.41 | 0.24 | 0.43 | 0.40 | 0.49 |
| Land | 0.64 | 0.48 | 0.78 | 0.42 | 0.81 | 0.39 | 0.59 | 0.49 |
| Self-employment | 0.56 | 0.50 | 0.61 | 0.49 | 0.54 | 0.50 | 0.56 | 0.50 |
| Professional | 0.27 | 0.44 | 0.22 | 0.42 | 0.11 | 0.32 | 0.30 | 0.46 |
| No. of observations | 2799 | | 381 | | 278 | | 2175 | |

Table 3.3 reports the same statistics for the subsample of households that report an occupation. This subsample is dominated by those who work so it is not surprising that the incomes and expenditures for these households are slightly larger than for the full sample. However, the comparisons between migrant and non-migrant households are qualitatively similar for the reduced sample as for the full sample.

In order to specify the partial, or marginal, impact that migration status of a household has on the differences we observe between log expenditure and log income, we require a

multivariate analysis that includes an array of variables that may influence this difference. We now turn to regression analysis, reporting these results in the next section.

### 3.4.2. Regression Results.

Table 3.4 reports the result of simple regressions with the log of the ratio of expenditures to income as the dependent variable and household characteristics as the independent variables. All models are estimated using ordinary least squares with the reported standard errors computed using 1000 bootstrapped replications.[12] Regression (1) is just the simple linear regression replicating the difference in means test between households with no current or recent migrants and households with current migrants and households with migrants who have recently returned (within the last 12 months). We see that households with current or recent migrants have significantly lower excess expenditure than households without any migrants. Households with a migrant who is currently abroad have excess expenditures that are 26.2% lower than the reference non-migrant households while households that have a recently returned migrant have excess expenditures that are 36.9% lower than the reference non-migration household. This supports our assertion that migrant income is a substitute for informal or non-reported activity. The full effect of the additional migrant income occurs after the migrant has returned but there is a significant impact even when the migrant is still abroad, most likely through remittances sent back to the household from abroad. This result is obtained using the full sample of households.

Regressions (2)-(4) report results for regression models that include the various household characteristics for the same sample. A number of important features are evident from these results. First, the coefficients on the two migrant indicator variables are consistent across

[12] Using bootstrapped standard errors allow for us to control for unobserved heteroskedasticity without the need to commit to the exact form or commit to the clustering variable needed to compute clustered-robust standard errors.

specifications and are always significant. The result that income from migrant labor is a substitute for informal or non-reported activity is robust to our different specifications.

**Table 3.4. Informal Sector and Migration Regression Results**

Dependent variable: log(expenditure/income)

| | Full Sample | | | | | | | | Working Sample | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variables | (1) | | (2) | | (3) | | (4) | | (5) | | (6) | |
| Migrant (abroad) | -0.262 | *** | -0.265 | *** | -0.260 | *** | -0.325 | *** | -0.278 | *** | -0.274 | *** |
| | (0.053) | | (0.053) | | (0.052) | | (0.050) | | (0.063) | | (0.063) | |
| Migrant (returned) | -0.369 | *** | -0.375 | *** | -0.369 | *** | -0.428 | *** | -0.617 | *** | -0.585 | *** |
| | (0.070) | | (0.070) | | (0.072) | | (0.071) | | (0.084) | | (0.080) | |
| Borrowing | | | 0.276 | *** | 0.279 | *** | 0.260 | *** | 0.264 | *** | 0.265 | *** |
| | | | (0.088) | | (0.087) | | (0.090) | | (0.097) | | (0.098) | |
| Vocational | | | 0.004 | | 0.022 | | 0.021 | | 0.083 | | 0.064 | |
| | | | (0.065) | | (0.063) | | (0.065) | | (0.070) | | (0.069) | |
| University | | | -0.001 | | 0.023 | | 0.060 | | 0.173 | *** | 0.047 | |
| | | | (0.047) | | (0.047) | | (0.051) | | (0.054) | | (0.061) | |
| Single | | | | | 0.009 | | 0.066 | | -0.145 | | -0.174 | |
| | | | | | (0.203) | | (0.192) | | (0.252) | | (0.231) | |
| no. of children(<15) | | | | | 0.015 | | 0.004 | | 0.003 | | 0.006 | |
| | | | | | (0.012) | | (0.012) | | (0.015) | | (0.015) | |
| no. of elderly (65+) | | | | | 0.121 | *** | 0.083 | ** | 0.081 | | 0.080 | |
| | | | | | (0.041) | | (0.039) | | (0.066) | | (0.066) | |
| Ethnic | | | | | | | -0.276 | *** | -0.292 | *** | -0.290 | *** |
| | | | | | | | (0.052) | | (0.055) | | (0.057) | |
| Urban | | | | | | | -0.124 | ** | -0.111 | | -0.119 | |
| | | | | | | | (0.063) | | (0.076) | | (0.077) | |
| Land | | | | | | | 0.294 | *** | 0.268 | *** | 0.270 | *** |
| | | | | | | | (0.064) | | (0.076) | | (0.076) | |
| Self-employed | | | | | | | -0.198 | *** | -0.036 | | 0.086 | |
| | | | | | | | (0.043) | | (0.048) | | (0.059) | |
| Professional | | | | | | | | | | | 0.349 | *** |
| | | | | | | | | | | | (0.073) | |
| Self-empl * Professional | | | | | | | | | | | -0.323 | *** |
| | | | | | | | | | | | (0.100) | |
| Constant | 1.047 | *** | 1.033 | *** | 0.955 | *** | 1.013 | *** | 0.812 | *** | 0.700 | *** |
| | (0.024) | | (0.028) | | (0.042) | | (0.075) | | (0.088) | | (0.093) | |
| Observations | 4391 | | 4391 | | 4391 | | 4391 | | 2799 | | 2799 | |
| $R^2$ | 0.012 | | 0.014 | | 0.017 | | 0.040 | | 0.051 | | 0.058 | |

Regression (2) adds variables that indicate a household's education level and whether or not they borrowed money in the past month. The coefficient on the variable *borrowed* is significant and positive with households who borrow having about 27% more excess expenditure than households without any borrowing. This number is consistent across the other specifications as well. This result is not surprising as in this dataset reported income does not include loans while it would be expected that expenditure would reflect the additional income due to loans. Including the borrowing dummy variable, however, does not affect our result that migration significantly decreases the amount of informal sector or non-reported income for households. The education variables are not significant and therefore do not appear to impact the excess expenditures of a household.

Regression (3) augments regression (2) with some household characteristic variables. The new variables that are included are an indicator variable on whether the household head is single and variables that indicate the number of young people in the house and the number of elderly people in the house. The marital status of the head of household and the number of young people under the age of 15 in the household do not significantly affect the excess expenditure for the household but the number of elderly does. This is not surprising as the pension paid to retirees in Tajikistan is very low and below the subsistence wage. Thus the elderly would need to augment their income. If this additional income was not reported in the formal income then we would expect to see an increased excess expenditure for the household as the augmented income would be likely to show up in the household expenditure.

Regression (4) adds in other household characteristics including the ethnicity of the head, whether the household is situated in an urban area, whether the household has access to cultivatable land, and whether any member of the household is self-employed. All of these variables significantly affect the excess expenditure of a household. Households with an ethnic

minority head have excess expenditures that are 27.6% lower than Tajik households. Households that are located in urban areas have excess expenditures that are 12.4% less than households in non-urban areas while households that have access to cultivatable land have increased excess expenditures of the order of 29.4%. Finally households with members who are self-employed have lower excess expenditures to the order of 19.8%. The result that ethnic minority households have significantly lower excess expenditures suggests that these households are not able to generate as much informal sector income as other households. The results for the urban households and households who have access to cultivatable land are consistent with each other. Households in rural areas have more opportunity to grow their own food which is included in the expenditures but not included in incomes.

The consistent result, however, is that households with migrants, have significantly lower excess expenditure than households without migrants. While not all of the discrepancy in reported expenditure over income can be attributed to informal sector activity, it is hard to believe that informal sector activity does not make up a large proportion of the discrepancy. Other sources of the reported discrepancy are likely due to unreported consumption of assets (animal stock, food, etc.) or to incorrect pricing of such activities. As reported in Table 3.2 and Table 3.3 the magnitudes in the differences in excess expenditure between households with migrants and households without migrants is large. It is very hard to believe that this difference is due to differences between households in pricing of in-kind consumption or consumption of assets. In particular it is hard to believe that households with migrants are systematically better (or worse) at pricing non-market activities than households without migrants to such an extent as to explain the large change in excess expenditure seen in the data. Therefore, the decline in the discrepancy between expenditures and income is likely to be due to the fact that remittances are explicitly measured in this survey while informal sector income is not explicitly measured. Thus our results

are consistent with the hypothesis that remittances from migrants are substitutes for informal sector income rather than complements.

Our regressions include controls for the other sources of income that are not included in the income data including loans and access to growing or rearing you own food. One source of additional income that we do not control for is additional income obtained by professionals "under-the-table." This income could not be considered informal sector income as the income is derived from activity that is identical to their formal sector income; the only difference is that it is not reported. In order to control for this we include a dummy variable for occupation but since households who are not working or unemployed do not report their occupation we include only working households in our sample. Regressions (5) and (6) are results using data from this restricted sample. Regression (5) is the same as Regression (4) with the only difference being the estimation sample while Regression (6) includes variables pertaining to a household's occupation; in particular a dummy variable if the household works in a professional occupation and the interaction between the professional dummy variable and the self-employment dummy variable.

The results are consistent with the first set of results that show that households with current or recently returned migrants having significantly lower excess expenditures than households without migrants. We also see that when restricting attention to only workers the significance of the elderly variable and the urban dummy variable disappears. This is not surprising in that we lose those households who are retired and not working and those households in the rural areas that are not working. The new result is that households whose head works in a professional occupation have significantly higher excess expenditure thus suggesting that there is additional income being collected "under-the-counter." This is reinforced by the result that professional but self-employed households do not show a significant increase in excess

expenditure. This is consistent with professionals who work in formal (and non–self-employed) jobs need to augment their income as their pay is low in their formal jobs.

The results reported above show that the discrepancy between reported expenditures and income for migrant households are significantly lower than those for non-migrant households. A possible criticism of the methods used would be that there are possible endogeneity biases present in our estimates that we have not modeled. However, in our specification the dependent variable is the amount of discrepancy between reported expenditure and reported income and it is not clear that excess expenditure over income would drive the decision to migrate. Certainly the total amount of expenditure (or income) might influence the migration decision but it is not clear that the component of expenditure that is unreported income would be a driver of the migration decision. Households would have to care about whether and from where their income was sourced for the excess expenditure over income to cause the migration decision.[13]

### *3.5. Conclusion*

Over the last 20 years, Tajikistan experienced both increasing migration and informal sector employment. They both are relatively new phenomena in Tajikistan, a former Soviet country, where informal sector employment and international migration were strictly controlled and even "prohibited" by the Soviet Government. After the Soviet Union's collapse, these restrictions quickly untwisted, now involving an appreciably large proportion of the population. Such preconditions make Tajikistan a good case to study, where one does not need to be very concerned about historically well-established patterns and traditions of migration and informality

[13]As a check, we have tried some possible instruments, including education variables, and find that the negative sign on migration is robust but that the magnitude is implausible large. The instruments that were tried had little theoretical motivation and were weak in the sense of having low first stage F-statistics. We therefore do not report the IV results for both of these reasons – i.e. the available instruments are weak and most likely not valid.

(or look further into the historical and cultural elements of these processes), allowing us to focus on economic issues and factors which help to explain how these two processes interact. Moreover, the very large size of the migration has made it a relative low cost path for obtaining additional income, and there are many households in the sample who have a current or recently returned migrant.

We consider informality and migration as alternative income sources for the household – the two are part of the portfolio of the family. When informality is considered in the context of migration, it is almost always in the context of the migrant working in an informal job, or not. In the household model we are implicitly considering in this paper, it is quite reasonable that one member of the household might migrate while another works in the informal sector. This is in line with the portfolio theory of migration, in which family members work in different labor markets as an income diversification strategy.

The Living Standards Survey allows us to investigate the relationship between external migration and local informal sector activity. To do so we used the discrepancy between reported household expenditure and reported household income as an indication of informal/unreported activity. We understand that all of this discrepancy is not due to only informal sector activity. However, variation across household in this discrepancy is most likely due to differences in informal sector activity – broadly defined – other sources for this discrepancy between expenditure and income, such as measurement errors and memory retention error, are not likely to differ systematically across households. Using this measure of discrepancy between expenditure and income we investigate the linkages between migration and the size of informal sector activity. We do this by estimating an equation that explains the discrepancy between expenditure and income using household characteristics and migration status.

The overall result that we find is that there is consistent evidence that migration (accompanied by remittances) and home region informal sector activities are largely substitutes for one another. This result is robust across all of our regression specifications. We find that households with members who have externally migrated are less likely to participate in the domestic informal sector in that they have a significantly lower discrepancy between their reported expenditure and reported income. We do not believe that there is anything special about migrant households that enable them to better measure or remember their expenditure or income than non-migrant households so that we conclude that the significant drop in the discrepancy between expenditure and income is due to the substitution of remittances, which are observed, for informal sector income which is not observed.

Our work indicates that the ability of professionals to engage in informal activities enables them to compensate for the discrepancy between their expenditure and income without migrating. Migrants typically are low-skilled non-professionals without post-secondary education who lack informal sector opportunities or might have lower earnings in the local informal sector. Migrants find it less costly to migrate (more earning opportunities are abroad) than to be involved in local informal sector. Migration becomes a substitute for informal sector employment.

We have documented the existence of this phenomenon and suggested some ways to understand its source. More work is needed – other case studies, modeling how and why this form of the link arises. The result adds a considerable amount of complexity to our understanding of the decisions faced by households in less-developed economies. Work on informality and work on migration should not continue to ignore the connection.

## Part III. Conclusion

In this dissertation we studied the effect of migration on three aspects of the economic life of migrants' relatives who remained in the source country. In the second chapter, we study the effect of migration on the job satisfaction of migrant's relatives who remain and work in the source country. According to our economic model, migrants' relatives build their expectations on earnings from migration through received information on the wage distribution in the destination country either from the size of remittances or directly from migrants. If their expected earnings from migration greatly exceed their current wages in the source country, migrant relatives become more dissatisfied with their jobs. Applying both parametric and semiparametric estimations to Tajikistan's data, as well as with controlling for an endogeneity issue with the variable of interest in the latter method of estimation, we found the significantly positive effect of the difference of the expected outside country earnings and current earnings of migrants' relatives on their job dissatisfaction. This result implies that a larger gap between what an individual could earn in the migration destination country and what she receives now at her current job in the source country makes that individual unhappier.

In our third chapter, we study effects of migration on decisions about acquiring professional education. We suggest controlling for both the size of remittances and migration's side effects. If the side effects of migration are omitted in the structural model, the estimate of the effects of remittances can be upward or downward biased, depending on whether the side effects favor or not an individual's decision on continuing education or not, respectively. To overcome this problem we use a dummy variable indicating whether there is a migrant the household. This way, holding remittances constant, the coefficient on the dummy variable captures the other side effects of migration as a change in the intercept between households with migrants and households without migrants. To deal with the multiple endogeneity issue, we used Bayesian

Limited Information Estimation based on applying Monte Carlo Markov Chain algorithms. For exclusion restrictions, we use the data on the current harvest and the prior to agricultural season migration. We found that remittances and the negative side effects of migration both have negative, but not statistically significant, causal effects on the decision of Tajikistan's young adults to acquire professional education. While ignoring the endogeneity issue, the estimation of the structural model shows significant negative correlation between the negative side effects of migration and the decision on professional education, and the coefficient on remittances is positive but not statistically different from zero.

In the fourth chapter, we look at how migration is related with the informal sector. The current literature defines them as being complementary, since new migrants may have difficulty in finding employment in formal work, so many of them end up informally employed. Tajikistan possesses both a very large informal sector and extensive international emigration. Using a newly introduced measure of the informal (unreported) activities as the gap between household expenditure and income, we find the significant and negative correlation between informal activities and migration: having a migrant member reduces the difference between household's expenditure and income. Furthermore, Tajikistan's professional workers ability to engage in informal activities enables them to forgo migration, while low-skilled non-professionals without post-secondary education choose to migrate instead of working in the informal sector. Our empirical evidence suggests migration and informality substitute for one another.

# Appendices

## *Appendix I: Limited Information Bayesian Estimation Model Experiment*

The econometric model is used from Tsurumi(1990) with some modification:

$$YA = XB + U,$$

where $A = \begin{pmatrix} 1.0 & -0.267 & -0.087 \\ -0.222 & 1.0 & 0 \\ 0.15 & -0.046 & 1.0 \end{pmatrix}$ and $B = \begin{pmatrix} 0.62 & 4.4 & 4.0 \\ 0 & 0.74 & 0 \\ 0.7 & 0 & 0.53 \\ 0 & 0 & 0.11 \\ 0.96 & 0.13 & 0 \\ 0 & 0 & 0.56 \\ 0.06 & 0 & 0 \end{pmatrix}$.

$X$ includes a constant, and six demeaned random variables drawn from the uniform distribution, $U$ is three vector variable drawn from the standard normal distribution. The variance-covariance matrix of $U$ is set to the following:

$$36\begin{pmatrix} 1.0 & \rho_\omega & 0.25\rho_\omega \\ \rho_\omega & 1.0 & 0 \\ 0.25\rho_\omega & 0 & 1.0 \end{pmatrix}, \text{ with } \rho_\omega = -0.90.$$

Two first endogenous variables, $y_1$ and $y_2$, were changed to dichotomous variables. The structural equation of interest is:

$$y_1 = 1\{0.222y_2 + 6.2x_1 + 0.7x_3 + 0.96x_5 + 0.06x_7 > u_1\}$$

The sample size is 1000. Number of replications is 1,100,000, where first 100,000 are removed, and then every 100th observation is kept.

Since two of three endogenous variables including the dependent variable in the structural equation were changed to binary, we can compare the estimated elasticities in the structural equation of $y_1$ to their true values:

**Table A.1. Model Experiment: Estimated Elasticities**

| Variables | True | Estimates | St.Dev. | AR(1) |
|---|---|---|---|---|
| $y_2$ | 0.1973 | 0.1694 | 0.0320 | 0.3823 |
| $y_3$ | -0.0972 | -0.0708 | 0.0334 | 0.3917 |

Estimates of elasticities of two endogenous explanatory variables in the structural equation are close to their true values with small variance and AR(1) coefficients.

## Acknowledgment of Previous Publications

The chapter entitles "Migration as a Substitute for Informal Activities" of this dissertation was published in co-authorship with Professor Ira Gang and Professor John Landon-lane as the separate article under the title “Migration as a Substitute for Informal Activities: Evidence from Tajikistan" in *Research in Labor Economics* (2012), Volume 34, in pages from 205-227. This paper applies the newly developed measure of informal activities developed in 2006 by Professor Ira Gang, Professor John Landon-Lane and Professor Ralitza Dimova in studying the impact of migration on informal sector.

The main responsibility of the dissertation's author in this joint research project was to work with the data from the 2007 World Bank's Living Standard Survey, to extract necessary information from the dataset, and to estimate econometric models of the impact of migration on the informal sector in Tajikistan. Using his knowledge on Tajikistan's economic development issues, he wrote parts of paper’s discussions on Tajikistan's economic background and its current issues related to the informal sector and the international migration.

**Bibliography**

Abdulloev, I., Gang, I., & Landon-Lane, J. (2012). Migration as a Substitute for Informal Activities: Evidence from Tajikistan. *Research In Labor Economics,34*(0), 205-227.

Acosta, P. (2006). Labor Supply, School Attendance, and Remittances from International Migration : the Case of El Salvador. *Policy Research Working Paper Series 3903, The World Bank.* doi:10.1596/1813-9450-3903.

Amuedo-Dorantes, C., & Pozo, S. (2006). Migration, Remittances, and Male and Female Employment Patterns. *American Economic Review , 96*(2), 222-226.

Amuedo-Dorantes, C., & Pozo, S. (2010). Accounting for Remittance and Migration Effects on Children's Schooling. *World Development, 38*(12). 1747‑1759. doi: 10.1016/j.worlddev. 2010.05.008

Angrist, D. (2010, February 8*). Mostly Harmless Econometrics: Multiple endogenous variables – now what?!.* Message posted to http://www.mostlyharmlesseconometrics.com/2010/02 /multiple-endogenous-variables-what-now/

Becker, G. S. (1991). *A Treatise on the Family* (Enlarged ed.). Cambridge, MA: Harvard University Press.

Blundell, R. W., & Powell, J. L. (2004). Endogeneity in Semiparametric Binary Response Models. *The Review of Economic Studies , 71*(3), 655-679.

Brown, R. P., & Leeves, G. (2007). Impacts of International Migration and Remittances on Source Country Household Incomes in Small Island States: Fiji and Tonga. *Working Papers 07-13, Agricultural and Development Economics Division of the Food and Agriculture Organization of the United Nations*. Retrieved from: ftp://ftp.fao.org/docrep/ fao/010/ah840e/ah840e.pdf

Cabegin, E. C. (2006). The Effect of Filipino Overseas Migration on the Non-Migrant Spouse's Market Participation and Labor Supply Behavior. *IZA Discussion Paper No. 2240*. Retrieved from: http://ftp.iza.org/dp2240.pdf

Calero, C., Bedi, A. S., & Sparrow, R. (2009). Remittances, Liquidity Constraints and Human Capital Investments in Ecuador. *World Development*, *37*(6), 1143-1154. doi: 10.1016/j.worlddev.2008.10.006

Card, D., Mas, A., Moretti, E., & Saez, E. (2012). Inequality at Work: The Effect of Peer Salaries on Job Satisfaction. *American Economic Review, 102*(6), 2981-3003. doi: 10.1257/aer.102.6.2981

Clark, A. E., & Oswald, A. J. (1996). Satisfaction and Comparison Income. *Journal of Public Economics , 61*(3), 359-381.

Cox Edwards, A., & Ureta, M. (2003). International Migration, Remittances, and Schooling: Evidence from El Salvador. *Journal of Development Economics , 72*(2), 429–461.

Deaton, A., & Muellbauer, J. (1981). Functional Forms for Labor Supply and Commodity Demands with and without Quantity Restrictions. *Econometrica , 49*(6), 1521-1532.

Diaz-Serrano, L., & Vieira, J. A. (2005). Low Pay, Higher Pay and Job Satisfaction within the European Union: Empirical Evidence from Fourteen Countries. *IZA Discussion Papers 1558*. Retrieved from: http://ftp.iza.org/dp1558.pdf

Dimova, R. & Gang, I. (2007). Self-Selection and Earnings during Volatile Transition. *Journal of Comparative Economics*, *35*(3), 612-629.

Dimova, R., Gang, I., & Landon-Lane, J. (2006). The Informal Sector during Crisis and Transition. In B. Guha-Khasnobis and R. Kanbur (eds) *Informal Labour Markets and Development* (pp. 88-108). Palgrave-McMillan Press.

Falkingham, J. (2004). Poverty, out-of-pocket payments and access to health care: evidence from Tajikistan, *Social Science and Medicine*, *58*(2), 247-258.

Fields, G. (1976). Labor Force Migration, Unemployment and Job Turnover, *The Review of Economics and Statistics*, *58*(4), 407-415.

Fields, G.(1975). Rural-urban migration, urban unemployment and underemployment, and job-search activity in LDCs, *Journal of Development Economics*, *2*(2), 165-187.

Fields, G.(1979). Place-to-Place Migration: Some New Evidence, *The Review of Economics and Statistics*, *61*(1), 21-32.

Gang, I. & Gangopadhyay, S. (1987a). Employment, output and the choice of techniques: The trade-off revisited, *Journal of Development Economics*, *25*(2), 321-327.

Gang, I. & Gangopadhyay, S. (1987b). Optimal Policies in a Dual Economy with Open Unemployment and Surplus Labour, *Oxford Economic Papers*, *39*(2), 378-87.

Gang, I. N. & Gangopadhyay, S. (1987c). Welfare Aspects of a Harris-Todaro Economy with Underemployment and Variable Prices. *The Developing Economies*, *25*(3), 203–219.

Gibson, J., McKenzie, D., & Stillman, S. (2010). Accounting for Selectivity and Duration-Dependent Heterogeneity when Estimating the Impact of Emigration on Incomes and Poverty in Sending Areas. *Policy Research Working Paper Series 5268, the World Bank*. doi:10.1596/1813-9450-5268

Grimm, M. & Gűnther, I. (2006). Inter and Intra-household linkages between the informal and formal sectors: a case study for urban Burkina Faso. In B. Guha-Khasnobis and R. Kanbur (eds) *Informal Labour Markets and Development* (pp. 155-180*)*, Palgrave-McMillan Press.

Grund, C., & Sliwka, D. (2001). The Impact of Wage Increases on Job Satisfaction - Empirical Evidence and Theoretical Implications. *IZA Discussion Paper No. 387*. Retrieved from: http://ftp.iza.org/dp387.pdf

Hamermesh, D. S. (1977). Economic Aspects of Job Satisfaction. In O. C. Ashendelter, & W. E. Oates, *Essays in Labor Market Analysis* (pp. 53-72). Jerusalem: Israel Universities Press.

Hamermesh, D. S. (2001). The Changing Distribution of Job Satisfaction. *The Journal of Human Resources , 36* (1), 1-30.

Harris, J. and Todaro, M. (1970), Migration, Unemployment and Development: A Two-Sector Analysis, *American Economic Review*, *60*(1), 126–142.

Ichimura, H. (1993). Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models. *Journal of Econometrics 5*(1-2), 71-120.

Ichimura, H., & Lee, L. F. (1991). Semiparametric least squares estimation of multiple index models: single equation estimation. In *Nonparametric and semiparametric methods in econometrics and statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics* (pp. 3-49). Cambridge: Cambridge University Press.

International Labor Organization (ILO). (1991). *The Dilemma of the Informal Sector, International Labour Conference, Report of the Director General*.

International Labor Organization (ILO). (2002). *Decent Work and the Informal Economy. Report VI for the 90th International Labour Conference, Geneva*.

International Labor Organization. (2010). *Migrant Remittances in Tajikistan.* Moscow: ILO Subregional Office for Eastern Europe and Central Asia, International Labour Office. Retrieved from:http://www.ilo.org/public/english/region/eurpro/moscow/info/publ /tajik_migr_remit_en.pdf

Killingsworth, M. R. (1983). *Labor Supply*. Cambridge: Cambridge University Press.

Kim, N. (2007). The Impact of Remittances on Labor Supply: the Case of Jamaica. *Policy Research Working Paper Series 4120, The World Bank.* doi:10.1596/1813-9450-4120

Kinyanjui, M. N. (2010). Social Relations and Associations in the Informal Sector in Kenya, *Social Policy and Development Programme Paper Number 43,* United Nations Research Institute for Social Development. Retrieved from: http://www.unrisd.org/publications/pp-kinyanjui

Klein, R. W., & Spady, R. H. (1993). An Efficient Semiparametric Estimator for Binary Response Models. *Econometrica, , 61* (2), 387-421.

Kristensen, N., & Westergaard-Nielsen, N. (2004). Does Low Job Satisfaction Lead to Job Mobility? *IZA Discussion Paper No. 1026*. Retrieved from: http://ftp.iza.org/dp1026.pdf

Lucas, R. (1978). On the size distribution of business firms, *Bell Journal of Economics*, *9*(2), 508-523.

McKenzie, D., & Rapoport, H. (2007). Network Effects and the Dynamics of Migration and Inequality: Theory and Evidence from Mexico. *Journal of Development Economics 84*(1), 1-24.

McKenzie, D., & Rapoport, H. (2010). Can Migration Reduce Eeducational Attainment? Evidence from Mexico. *Journal of Population Economics 24*(4), 1331-1358.

McKenzie, D., & Sasin, M. (2007). *Migration, remittances, poverty and human capital: conceptual and empirical challenges.* Washington, DC: World Bank. Policy Research Working Paper No. 4272. Retrieved from: https://openknowledge.worldbank.org /bitstream/handle/10986/7453/wps4272.pdf?sequence=1

Mincer, J. (1970). The Distribution of Labor Incomes: A Survey With Special Reference to the Human Capital Approach. *Journal of Economic Literature , 8*(1), 1-26.

National Bank of Tajikistan. (2011). *National Bank of Tajikistan.* Retrieved September 17, 2011, from Statistics of Real Sector: http://nbt.tj/ru/statistics/real_sector/

Nguyen, T., & Purnamasari, R. (2011). Impacts of International Migration and Remittances on Child Outcomes and Labor Supply in Indonesia: How Does Gender Matter? *Policy Research Working Paper Series, No 5591, The World Bank*. doi:10.1596/1813-9450-5591

Noda, S. (1999, March). Aftermath of War in Tajikistan: Country Profile. *International Labour Organization News Eastern Europe and Central Asia, 3*. Retrived from: http://www.ilo.org/public/english/region/eurpro/moscow/info/files/99_1en.pdf

Olimov, J. 2007. *United Nations Human Development Report 2007: Informal Economy in Tajikistan*, Dushanbe: United Nations Development Programme.

Perry, G., Maloney, W., Arias, O., Fajnzylber, P., Mason, A., and Saavedra-Chanduvi, J. (2007). *Informality: Exit and Exclusion.* Washington, DC: World Bank. Retrived from: https://openknowledge.worldbank.org/bitstream/handle/10986/6730/400080Informal101 OFFICIAL0USE0ONLY1.pdf?sequence=1

Rivers, D., & Vuong, Q. H. (1988). Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models. *Journal of Econometrics*, *39*(3), 347-366.

Robertson, L. R. (ed.). (1999) *Russian & Eurasia Facts & Figures Annual*, *25*(2), Academic International Press.

Rodriguez, E., & Tiongson, E. (2001). Temporary Migration Overseas and Household Labor Supply: Evidence from Urban Philippines. *International Migration Review*, *35*(3), 709-725.

Rothe, C. (2009). Semiparametric Estimation of Binary Response Models with Endogenous Regressors. *Journal of Econometrics, 153*(3), 51-64.

Russian State Statistical Committee. (2010, December 1). *Data*. Retrieved December 12, 2010, from Russian State Statistical Committee: http://www.gks.ru/wps/PA_1_0_S5/Documents/jsp/Detail_default.jsp?category=1112178611292&elementId=1140086922125

Schroeder, G. E. (1981). Regional Living Standards. In I. Koropeckyj, & G. E. Schroeder, *Economics of Sovient regions* (pp. 118-156). New York: Praeger.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

Staiger, D., & Stock, J. (1997). Instrumental Variables Regression With Weak Instruments. *Econometrica, 65*(3), 557-86.

Statistical Agency of Tajikistan. (2010, December 1). *Real Sector Data*. Retrieved December 12, 2010, from Statistical Agency of Tajikistan: http://stat.tj/en/database/real-sector

Statistical Agency of Tajikistan. (2011). *Database*. Retrieved September 17, 2011, from Statistical Agency of Tajikistan: http://stat.tj/ru/database/real-sector/

Statistical Committee of CIS. (2011). *Average Monthly Nominal Wage in the CIS Countries, in national currency*. Retrieved September 17, 2011, from Statistical Committee of CIS: http://www.cisstat.com/index.html

Stock, J. H., & Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In J. Stock, & D. Andrews, *Identification and Inference for Econometric Models: Essays in Honor of Thomas J Rothenberg* (p. 80-108). New York: Cambridge University Press.

Thorbecke, E.(1999). A dual-dual framework to analyze intersectoral relationships throughout the development process. In G. Ranis, Y-P Chu and S-C Hu, (Eds.) *The Political Economy of Comparative Development Into the 21st Century*, Edward Elgar Publishing, Ltd.

Tsurumi, H. (1990). Comparing Bayesian and non-Bayesian limited information estimators. In S. Geisser, J. Hodges, S. Press, & A. Zellner, *Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard* (pp. 179-207). Amsterdam: North-Holland.

Tybout, J. (2000). Manufacturing firms in developing countries: how well they do, and why? *Journal of Economic Literature, 38*(1), 11-44.

United Nations Development Programme and the Center for Strategic Studies under the President of the Republic of Tajikistan. (2010), *Corruption in Tajikistan - Public Opinion*. Retrieved October 17, 2011, from UNDP: http://www.undp.tj/files/undpeng.pdf

United Nations. (2011). *Population below national poverty line, total, percentage*. Retrieved September 17, 2011, http://data.un.org/Data.aspx?d=MDG&f=seriesRowID%3A581

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data.* Cambridge; London: The MIT Press.

World Bank. (2008, November 1). *Russian Economic Report*. Retrieved December 12, 2010, from World Bank: http://siteresources.worldbank.org/INTRUSSIANFEDERATION/Resources/rer17_eng.pdf

World Bank. (2011). *Migration and Remittances. Factbook 2011. Second edition.* Retrieved August 12, 2011, from World Bank: http://siteresources.worldbank.org/INTLAC/Resources/Factbook2011-Ebook.pdf

World Bank.(2011). *World Data Bank*. Retrieved September 17, 2011, from World Bank: http://databank.worldbank.org/ddp/home.do?Step=12&id=4&CNO=2

Yang, D. (2008), International Migration, Remittances and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks, *The Economic Journal*, *118*(528), 591–630. doi: 10.1111/j.1468-0297.2008.02134.x