

UCL
Université
catholique
de Louvain



**UNIVERSITÉ
DE GENÈVE**

GENEVA SCHOOL OF ECONOMICS
AND MANAGEMENT

Polarisation of skills in developing countries: a case study of India

Thesis presented by
Stéphane Berberat

Supervisor
Florian Mayneris

Reader
Frédéric Docquier

Academic year 2016/2017

In order to obtain the Double Degree
Master 120 en Sciences économiques, Orientation générale, Finalité spécialisée (UCL/UNamur)
and
Master of Science in Economics (UNIGE)

Acknowledgements

I would first like to thank my thesis promotor Professor Florian Mayneris at the Université catholique de Louvain. Professor Mayneris was always available whenever I had a question about my research or writing throughout the development of this master's thesis. The pages that follow owe a lot to his shrewd guidance, and I will be forever grateful for helping me complete this work.

I would also like to acknowledge Professor Frédéric Docquier of the Economic School of Louvain at the Université catholique de Louvain as the reader of this thesis. I am greatly indebted to him for introducing me to the IPUMS INTERNATIONAL database as well as for all his support along the way.

I wish to acknowledge the statistical offices that provided the underlying data making this research possible, namely the Ministry of Statistics and Programme Implementation of India.

I would like to especially thank Simon Bouvier, my best friend, for proofreading my thesis and supporting me during the whole process of putting it together.

I would like to thank Elena Salmaso for proofreading my thesis and helping me with the structure of it, and Sayali Javadekar for her precious help concerning India's education system and all the precious information she provided me with advice on Indian culture.

My sincere thanks also go to Joris Colaco for his support, perspective and suggestions relative to gender studies and political science, as well as Angela Doku, Raoul Theler and Jonathan Teixeira for their suggestions and help.

Finally, I must express my very profound gratitude to my parents and for providing me with support and encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Table of content

| | |
|---|----|
| 1. Introduction..... | 1 |
| 2. Methodology..... | 5 |
| 3. Data..... | 7 |
| 4. Results..... | 11 |
| 4.1 Descriptive statistics..... | 11 |
| 4.2 Mincer equation..... | 17 |
| 4.2.1 Mincer equation in its simplest form as a building block..... | 18 |
| 4.2.2 Interactions between migration, gender and urban area..... | 20 |
| 4.2.3 Reasons for migration and correlations of occupations..... | 23 |
| 5. Concluding remarks and discussion..... | 28 |
| 5.1 Limitations..... | 30 |
| 5.2 Related field and further research..... | 30 |
| 5.3 Conclusion..... | 31 |
| References..... | 32 |
| Appendix..... | 35 |

List of figures and tables

| | |
|--|----|
| Figure 1 Wage distributions for migrants and non-migrants | 11 |
| Table 1: Shares of reason for migration regarding the previous location pooled | 12 |
| Table 2: Reasons for migration by previous area of living pooled | 13 |
| Figure 2 Migration reasons by gender..... | 14 |
| Figure 3 Wage distributions for migrants and non-migrants according to the gender | 15 |
| Figure 4 Occupation by gender for non-migrants and migrants | 16 |
| Table 3 Mincer equations in its simplest form | 19 |
| Table 4 Mincer equations augmented and controlled by year fixed effect | 20 |
| Table 5 Simple interaction and triple interaction between migrants gender and urban | 22 |
| Table 6 Interaction between urban area and migrants for working reason | 24 |
| Table A.1 Shares of reason for migration regarding the non remunerated agents | 35 |
| Table A.2 Wage distribution statistics by area of living | 35 |
| Table A.3.1 Two-sample Kolmogorov-Smirnov test for equality of distribution between rural and urban area | 35 |
| Table A.3.2 Two-sample Kolmogorov-Smirnov test for equality of distribution between genders..... | 35 |
| Table A.4 Wage distribution statistics by gender..... | 36 |
| Table A.5 Years of schooling versus year proxy | 36 |

1. Introduction

Internal migrations in developing countries mainly concern rural to urban or rural to rural migrations. These migration patterns are predominant in China, India and several Sub-Saharan African countries, as opposed to Latin American migrations for which the pattern is urban to urban.

In addition to their pattern, migrations also have their set of constraints. These are mainly financial but can also relate to distance and lack of information. These constraints usually limit rural to urban migration but support rural to rural. Regardless of the particular pattern followed or constraints, the vast majority of migrants end up being better off. Migrations have reduced poverty in China, India and to some extent in African countries, albeit indirectly thanks to migrants' remittances to relatives left in their village of origin (Lucas (2015)).

Specifically in India, internal migration is a well spread phenomenon, in large part thanks to a well-connected railroad network allowing easy long distance travel. The network has been greatly improved since 2012 with the construction of the Golden Quadrilateral, a highway network connecting India's main cities. Most reasons for migration—from labour to marriage or even education—involve agents either attracted by the benefits that urban areas might offer, or attempting to extricate themselves from the quagmire of poverty. Although a well-connected network of transportation alleviates the cost of migration, it remains a costly decision not only monetarily, but also psychologically (Bhagat 2010; Czaika 2012; Lucas 2015). For example, migration places couples on the horns of a dilemma; either the couple separates or adapts. The latter is a common feature in India.

On this aspect Banarjee (1981) explained internal migration in India in the context of family. Furthering the analysis of Mincer (1978), where migration is assumed to be a joint decision of a nuclear family, the author adapts the analysis for India, where the migration involving the whole family is rare. One of the main findings is that migration seeks to maximise benefits for families rather than individuals. Despite the distance, links between family members stay strong. Husbands either remit to their spouse or visit them in the village of origin. De Haan (1997) speaks of migration as being circular rather than permanent: migrants return to their home once they have saved enough money for their needs. However, Banarjee concludes his article inviting further research on migration, such as including reasons beyond labour, and expanding analysis to include women as well as men. Despite being attractive, migration remains highly affected by discrimination.

In addition to the high costs of migration, migrants suffer from a segmented labour market (De Haan (1997)). Indeed women, the poor and lower caste members endure severe discrimination and are often made scapegoats for the woes of cities, (Bhagat (2010) and Khan (2016)). This discrimination is not only confined to urban areas but is also present in rural ones, where women live in constant fear of

sexual harassment. In order to prevent or at least to attenuate this problem some men from the same village work with them in the fields (Rao and Rana (1997)). Despite the challenges that migrations present, migrants are better off in terms of income than non-migrants, even if the difference is small (De Haan (1997)).

The aim of this paper is to investigate the heterogeneity of income premia to migrants along several dimensions such as gender, reasons for migration and levels of skill by using the IPUMS-International database on India, for the years 1983, 1987 and 1999. A general note that we want to stress in this paper is that all the links being analysed do not aim to determine causality between the set of explanatory variables and the income of migrants, but rather to clarify the relationship among these variables.

Our empirical strategy to analyse the relationship was carried out in two steps: firstly, we investigated the descriptive statistics in order to assess how the distribution of wages between migrants and non-migrants relate, as well as the distribution between women and men. Then we pushed our analysis a step further by running a Mincer equation quantifying a premium for observable characteristics. Thereupon we created a more complex model by introducing simple and three-way interaction terms to assess the heterogeneity of the premium along several dimensions (urban areas occupations, etc.)

Our results show that the “urban premium” benefits male migrants, regardless of their reasons for migration. Women who migrated suffer from their gender status but migrations toward cities alleviate the penalty on their income. These results have been confirmed by Khan (2016). The premium benefits migrants going to an urban area. However the author underlines that the premium is relative to gender, caste and tribe, which remain severe discrimination factors in the Indian labour market.

The most important references for this paper are the articles of Khan (2016) and Bhagat (2010). The contribution of this paper to the existing literature is that it bridges these two articles. Indeed, the former evaluates the wage differential between migrants and non-migrants carrying out an analysis to some extent similar to the one in this paper. By using the National Survey Sample (NSS) for India relative to the years 1999-2000 and 2007-2008 rather than the Census, Khan explains the wage differential between migrants and non-migrants. His analysis starts with a Mincer equation and some key control variables in order to understand the main features of these categories. Then he pushes the analysis further by using a RIF-quantile regression decomposition to gauge the wage gap more precisely across the distribution of wages.

The second main reference of this paper is the article by Bhagat (2010), which is more descriptive and explains the different patterns and reasons for migration between men and women. We also further its analysis by quantifying a premium on income for observable characteristics among the different

reasons for migration. Moreover, we also explore the heterogeneity in premiums between different levels of skills. We found that the skill level of migrants brings an important premium especially in cities but the reasons for migration matter a lot: the larger premia are not observed among those who migrate for work, but other categories, mainly those who migrate for educational reasons.

This paper contributes to the literature by taking wage differential analysis a step further, focusing on heterogeneity of premia among migrants and non-migrants according to their areas of living (rural or urban) and their reasons for migration. Some features of current literature explain that men and women follow very specific migration patterns. The former are described as mainly migrating toward cities due to work reasons and across state lines. Specifically on India, De Haan (1997) relates that migrations mainly concern men in search of work because of lack of income or land. Conversely to what we might assume, De Haan emphasizes that migration is not necessarily predominant among the poorest. Bhagat (2010) stressing that migration is not only seen among poor agents but rather the opposite; agents with some background skills and education seem to fit the profile of the average migrant better. Indeed the former have nothing to lose while the latter can reap the benefit of urban migration (Mahapatro (2014)).

More specifically on women's migration, it is known that they are bitterly discriminated against in the labour market (Rao and Rana (1997)). Lusome and Bhagat (2006) relate the trend and patterns of migration in India explaining that internal migration is a widespread phenomenon, which concerns about a third of the population. By using data from Indian censuses, they describe the magnitude and the different reasons for migration by gender from 1971 to 2001. They observe specific patterns of migration for women and men; the former migrate especially for weddings while the latter do so for work. Identical patterns are seen for the years after 2001 (Bhagat (2010) Abbas and Varma (2014) and Khan (2016)). Despite being shown very clear differences in reasons for migration between women and men, Rao and Rana (1997) assess that women migration is not only a matter of marriage, since husbands often leave their spouse behind once they migrate for work. Being in charge of the children and on their own, they have to adapt to avoid starvation. Women are traditionally seen gathering forest goods to bring additional income and/or to feed the household. Migration has been exacerbated by deforestation, pushing women to migrate in order to find other sources of revenue

Our second contribution is to assess the wage differential between men and women, carrying an analogical analysis between migrants and non-migrants according to their gender.

Much of the literature states that women tend to travel short distances, staying within the state or even in the same district, leaving their family household only to join their husbands once they get married (Abbas and Varma (2014) and Bhagat (2010)). Concerning the destination of migration, Lusome and Bhagat (2006) explain that before the nineties, migration flows were essentially concentrated within

states, while interstate streams were quite weak. Czaika (2012) adds that the destination of migrants depends on relative and absolute deprivation of household (Stark 1984; Stark and Yitzhaki 1988): migrants seek to improve their relative position with respect to a reference group (i.e. overall population, specific social group). He distinguishes between relative deprivation related to short distance migrations, within the states, and absolute deprivation concerning long distance and international migrations. Actually, male migration is often further, interstate and longer (about six months) compared to women's short commuting moves, which last about a month. Srivastava and Sasikumar (2003) underline that intradistrict migration concerns especially women, who leave their parental home to move to their husband's household. Intrastate migrations are related to rural-urban moves while shorter ones are usually rural to rural ones. (Lusome and Bhagat 2006; Keshri and Bhagat 2013). The rural to urban stream became stronger and stronger in the nineties with economic liberalisation and is expected to extend, while the reverse to rural to rural flow is becoming weaker (Lusome and Bhagat 2006; Bhagat 2010).

Finally, we also contribute to the literature on the relationship between occupation and wages according to different levels of skill. On this subject, Chandrasekhar and Sharma (2014) stress another interesting feature of migration: migrants are looking for the best opportunities in terms of education and income to fit with their qualifications. The authors point out that migrants for education are likely to migrate to more developed states, if not abroad (what is known as "brain drain") in order to benefit from higher education quality, infrastructure and future higher income. Hence the authors highlight the pressing need for investment in education and particularly post-secondary education in the interest of economic development.

The rest of the paper is organized in four sections. Section 2 presents the methodology, describes the IPUMS-International dataset and the principal variables of interest. Section 3 describes the IPUMS-International dataset and the principal variables of interest. Section 4 presents the results of the descriptive analysis and the Mincer regression on the log wage by migrant characteristics. Section 5 discusses the results, widens the research to urban economics and concludes.

2. Methodology

On the basis of the theoretical insights and on previous empirical findings presented so far, we investigated several dimensions of the income premium: firstly the urban and rural aspect, then the situation regarding women's migration and, finally, the motivation for migrations.

We were expecting male migrants to benefit from a premium in cities. As Lucas (2015) underlines, migration especially concerns workers. However, we supposed that other reasons for migration also matter. Indeed migration for weddings and toward cities might concern women from the upper sphere of society, or migrants for education, who might benefit from the premium offered by urban areas as well. To some extent, we expected our findings to be similar to Khan's (2016) but through a decomposition of migrant categories, allowing us to observe whether working migrants are the ones reaping the most benefits from living in cities compared to all the other migrant categories.

We have seen through the literature review above that women are severely discriminated against as migrants. Mainly migrating to short distance destinations, they suffer from their gender status as Rao and Rana (1997) point out. However, we supposed that even if women are associated with wedding migration, this is not women's only reason for migrating, and more investigation was required. Despite being aware of how the highly segmented Indian labour market works against them, women have had to go beyond the traditional means of raising additional income in their village of origin. Migrations toward cities might be a way to do just that. If that were to be the case, the premium from migrating in urban areas would attenuate the negative impact of discriminations without cancelling it.

We expected the more skilled agents to migrate to urban areas as Chandrasekhar and Sharma (2014) found. Migrants for education are looking for the best opportunities to work or get a tertiary education. Hence, we were expecting larger premiums in cities either for those working after their studies or for those who are highly skilled.

In this paper we carried out our analysis on migrations and their associations on wage by estimating Mincer equations. The raw form of this equation simply seeks to explain the wage relation between the education and the professional experience of an individual using ordinary least square regressions (OLS). These regressions have been under fire from critics for three reasons: endogeneity problems, non-linear education premiums and homogenous experience as related by Humphreys (2013).

To tackle the first problem we could have either tried to find instruments to solve the endogeneity or introduced key control variables. The downside of the former was that it entailed finding instruments good enough to break endogeneity but also independent of other latent variables affecting wage, all this without triggering other problems. (Card, D., 1999: 1819-1822). Therefore, we chose to introduce

key control variables. However, it is important to keep in mind that including a very extensive list of variables in the Z vector (see equation (2) below) will never be exhaustive (Harmon et al. 2003: 138).

Another issue to which we were confronted was the absence of an experience variable. As a lot of datasets do not have information on this variable, we approximated it thanks to the following expression –the Mincerian experience. (Cahuc, P. and Zylberberg, A, 2004: 85-87).

$$Experience = age - (years\ of\ schooling + 6) \quad (1)$$

It is worth stressing that the coefficient 6 in equation (1) refers to age at the beginning of schooling (Mincer, 1974: 84). However, the application of this equation is dependant on the availability of data on years of schooling.

In our sample, we only had years of schooling observations for 1983. To solve this issue we followed a method developed by Heckman et al. (2003: 48-49) using Census data: instead of years of schooling, the highest-grade achieved in school was used. As our sample contained observations on the grades obtained by individuals, we chose to follow this methodology. Indeed, we used the variable labelled as *years of education* in the following section. We impose a certain number of years of education for each degree achieved, from 0 year if primary school has not been completed to 15 years for the achievement of a university degree.

The equation below was benchmark for our analysis. As we used a Mincerian equation, taking care to discard the age variable from the equation (2) in order to avoid collinearity with the experience variable.¹

$$\ln(Y) = \beta_0 + \rho s + \beta_1 x + \beta_2 x^2 + \beta_3 Z + \varepsilon \quad (2)$$

Where $\ln(Y)$ is the logarithm of the wage, β_0 is the constant, s the years of schooling, x the years of potential experience, x^2 the experience squared and Z a vector of other variables.

The starting point of the study was to focus on equation (2), where we controlled for gender, urban area of living occupation, religion and marital status. The very first equation takes into account only year fixed effect. Indeed the analysis was first carried out keeping the years fixed, and later controlling for year-occupation fixed effect, meaning that we not fixed the years anymore but also fixed the same occupation categories. In order to assess the premium or penalty of the different variables we also considered their interactions, firstly, on a two-way interaction, then on a three-way to be even more precise.

¹ Summary descriptive statistics can be seen in table A.2

3. Data

Database description

We wanted to explain the wage of migrants depending on their reasons for migration and according to individual characteristics. Therefore, we needed the data to report not only migration information but also labour market and socioeconomic data points for workers.

The Integrated Public Use Microdata Series (IPUMS) database satisfies these conditions. These data were provided by the Minnesota Population Center and reflect data from censuses and surveys worldwide². As we are studying India our sample relates to the Census reported by the Indian Ministry of Statistics and Programme Implementation. To create our sample we proceeded in two steps. Firstly, we had to choose the sample we wanted to work with (country and available years) and second, we had to select the variables of interest split into two main categories:

Household: containing subcategories of variables such as household composition, household economics or geographical variables.

Person: containing information on family structure, sociological variables (gender, age marital status...), labour market variables and migration.

In order to allow comparability across the years we decided to choose the year where the variables on migration are available. Thus we ended up with a cross-sectional sample composed of three years, 1983, 1987 and 1999. Our paper covers 87061 workers, 50410 and 85523, for each year mentioned, respectively. As the paper proceeds, we refer to these numbers where migrants and non-migrants workers are pooled. This being a multiple cross sections database, we were not able to follow migration flows by observing the agent through time. Nonetheless, we were able to study the reasons for migration and the previous state of residence for agents, since state of origin data is available for every observation in our database. The former can be studied thanks to a variable relating the reasons for migration discussed below, and the latter explains the most recent moves of the individual between administrative units.

Migrants

First of all, we would like to emphasize some problems with the measurement of our sample. As described above, our database is the fruit of harmonised work on censuses and surveys. It should be pointed out that the definition of a migrant is understood in different ways in different datasets. For

² For further information on IPUMS International we warmly encourage the readers to consult the following link: <https://international.ipums.org/international-action/faq>

instance, in the National Sample Survey (1981: 5), migrants are defined as “a normally resident member of a sample household was treated as a migrant if the person’s village/town of enumeration was different from his/her last usual place of residence (village/town)”. However, the Census definition is even wider: “A person whose birth place/last residence is different from the village or town of enumeration.” (Census of India, 1981: 210).

Activity status and age of working

In our paper, we focus exclusively on the subsample composed by the population in age of working and being employed. This means the employed population between 15 and 64 years of age.

We justify this choice by the fact that this age group represents between two thirds and three quarters of the whole population for the years studied once agents below 15 and above 64 are taken out.

Income

The remuneration of the agent is reported on the year of the census. The data is for weekly wages, paid in the local currency (Indian Rupee) and not inflation-adjusted. As wage distributions are often positively skewed, taking the natural logarithm of the wage might be a good idea to have a better distribution. Moreover, using a logged variable as dependant variable allows us to carry a semi-elasticity interpretation if the independent variables are level.

However, taking the log of the wage eliminates all the observations on workers who work but are not remunerated. As our analysis is focused on people in age of working who are employed *and remunerated*, we have to be more precise. As non-remunerated workers compose more than the half of our sample, we chose to dispose of them in order to have a normal shape of the log wage. Non-remunerated workers are those who declared, during the survey, a received or receivable income of 0 rupees on a weekly basis. Excluding these individuals facilitated the interpretation of the subsequent analysis on wage.³

Distance proxy and migration categorisation

The choice of the IPUMS-International database on India was mainly driven by its extraordinary information on migration. Indeed, thanks to three variables on migration we were able to know the reasons for people movements, the previous state of residence and the migration status according to the previous residence. The latter informs us on the nature of migration from a geographic point of

³ Table A.1 shows the share of the non-remunerated agent by gender. We realise that 89.75% of men and 43.67% of women are non-migrants. Together the non-migrants category accounts for 78% of the subsample. The migration concerning women is especially for wedding reasons.

view, split in five categories: non-migrants, intra district migration, intrastate migration, interstate migration or migration from abroad. Thus we are able to know from which states individual come from but not the destination. Moreover, we were able to use this variable relating the reasons for migration, as a proxy for distance following Gobillon's (2001) approach for reasons for migration, as featured on the tables in section 4.

On the subject of migration categorisations, it should be noted that we chose to follow a classification similar to the one used by the United Nation Population Division (1970: 64-65) because of the many reasons for migration. We also decided to create four categories of migration in order to facilitate our analysis: the first being working reasons composed of migrants in search of jobs, migrants whose jobs were subject to relocation, migrants who were assigned a job assignment away, or migrants in search of opportunity. The second is relative to family reasons and concerns mainly family moves but not weddings. Indeed, since weddings are the main reason for migration of women in India, we decided to create a full-fledged category for wedding migration. Finally, the last category relates to other reasons for migration composed by migration for studying, political and social problems and other unclassified reasons.

Repartition of skills

Usually skills are approximated by education but in our sample we have a variable reporting the category of job occupation. This variable uses the International Standard Classification of Occupations of 1988 (ISCO-88), allowing us to divide abilities according to a scale of skills:

The first level concerns the elementary occupations, mostly simple and repetitive tasks, which can be very physical.

The second level is quite heterogeneous. It is composed of occupations such as plant and machine operators, craftspeople and related trades, skilled agricultural and fishery workers, service workers and shop and market sales workers, and, finally, clerks.

The third level is for technicians and associate professionals. These occupations focus on technical work related to specific field such as life science and social sciences. It can also involve teaching at certain educational levels.

The fourth level, professionals, is very similar to the third level, but it goes further insofar as the main task of migrants in this category is the production of knowledge, teaching, and and undertaking research in a methodical manner.

The last category concerns legislators, senior officials and managers. This category doesn't make any reference to levels of skill. The main occupations are focused on formulating government policies,

laws and public regulations. Tasks involving acting as a representative, and coordinating firms or departments are also included (ISCO-88, 2004).⁴

We decided to gather the third level, the fourth level and the legislators and managers category to form one single high skill occupation group. Despite the absence of skill level references in their category, the legislators and managers category can be easily associated with the fourth level of skill (ISCO-08)⁵. We grouped ISCO level two occupations together as intermediate skill occupations, while elementary occupations are classified as low skill.

Urban areas

As a determinant variable of our paper is the place of living of the individuals, it is important to have a clear definition of what constitutes an urban area. We followed IPUMS' definition of such areas: an urban area is defined as town from an administrative perspective, having an authority entity such as a municipal corporation, having at least 5000 inhabitants or 400 per square kilometre and "pronounced urban characteristics and at least three fourths of the adult male population employed in pursuits other than agriculture."⁶

Years of education

In India, education follows a 10+2+3 pattern, which means ten years of mandatory school to reach the secondary level, two of high secondary and 3 for a bachelor's degree (4 if it is a technical degree such as engineering). Our dataset allows us to decompose the mandatory years of schooling in primary, middle and secondary grade. That means we were able to build a proxy for years of schooling using the following method: if primary education has not been completed, the years of schooling take the value of 0; having completed the primary school is equivalent to 5 years of schooling; completing middle school completed is equal to 8 years while finishing secondary school is worth 10. Finally, as higher secondary is a kind of preparation for taking an undergraduate degree, we chose to follow the initial variable repartition and attribute 15 years of schooling for the accomplishment of a university degree. (British Council, 2014: 12-13)

⁴ Detailed information on skill categories are available on the ILO website: <http://www.ilo.org/public/english/bureau/stat/isco/isco88/publ4.htm>

⁵ The index correspondence (part 4) provided on the link below, allows us associate the legislators and managers group to a fourth skill level category. <http://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm>

⁶ https://international.ipums.org/international-action/variables/MARST#description_section

4. Results

Firstly, we focused on the descriptive statistics of our subsample to get a sense of the disparities between migrants and non-migrants with regards to the wage premia and the reasons for migrations. Then we examined the remuneration of migrants with the Mincer equation in its simplest form. This analysis was then expanded into more precise stages, during which we added a set of controlling variables such as gender, levels of skill and even marital status. Subsequently, we also examined the interactions between urban areas and migration for working reasons, as well as the link between occupation and migration for working reasons. We then topped off our analysis with triple interactions, wherein occupation, reason for migration and urban area were studied altogether.

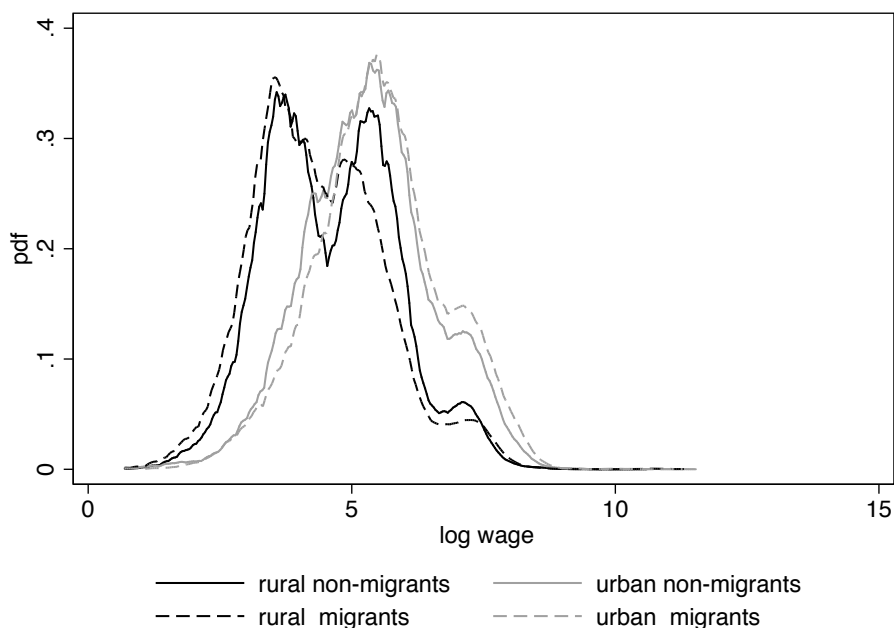
4.1 Descriptive statistics

At the beginning we wanted to determine if there were any disparities between migrants and non-migrants with respect to the areas where they live.

Figure 1 shows that migrants living in rural areas (black dashed line) seem to have been less remunerated than non-migrant workers also living in rural areas (solid black line).

Meanwhile in urban areas the opposite seems to occur: migrants were paid better than workers who already lived there. Indeed, the distribution for migrants living in cities (dashed grey) was slightly more negatively skewed than the non-migrants one (solid grey).

Figure 1 Wage distributions for migrants and non-migrants



The characteristics of the distributions in table A.1 of the appendix, show that the mean of the log wages is slightly larger for those who migrated in urban areas and slightly smaller for those who migrated to rural areas. Interestingly, similar distributions of wages between migrants and non-migrants are shown in Khan (2016). To ensure that the distributions of Figure 1 are not equal, we followed Khan’s methodology and ran a Kolmogorov-Smirnov test⁷. Indeed, the test showed migrants living in cities benefitted from a premium, while those living in rural area suffered from a penalty. A possible explanation that the author gives is the over-representation of women in rural areas, which pushes the overall wages of migrants down. (Khan I., 2016: 252-259). However, our data does not make the case for this explanation as clear as his paper makes it out to be.

We then wondered how the distribution of migration reasons relates to migration distances. In order to answer this question, we concentrated on our sample of 83216 migrants. As table 1 shows, the reasons for migration variable can be expressed as a function of the distance travelled between the previous location and the current one. The shortest distance, migration within districts, is the most common. Further distances decreased migration dramatically. Inter states migrations stayed relatively high, reaching 29.22% for working migration, while wedding is usually settled within districts, with 78.06% of wedding migrants coming from the same district as the previous location. Migrations from abroad were minimal.

Table 1: Shares of reason for migration regarding the previous location pooled

| | Intra district | Inter district | Interstates | Abroad | |
|----------------|-----------------------|-----------------------|--------------------|---------------|-------|
| Working | 35.48 | 34.24 | 29.22 | 1.05 | 38997 |
| Family reasons | 44.08 | 29.73 | 21.09 | 5.11 | 8873 |
| Wedding | 78.06 | 16.66 | 5.1 | 0.18 | 25932 |
| Others | 57.23 | 22.76 | 12.55 | 7.43 | 9414 |
| Observations | | | | | 83216 |

Note: The Table represents the percentage of migration by categories. As explained in section 3, the migrations across space allow us to know the provenance of the migrants but not the destination.

⁷ We can strongly reject the null under which the distributions are equal for figure 1. The outcome of these tests are shown in table A.3.1 in the appendix

These patterns of migration are directly in line with the results of Gobillon (2001) about human movement in France, where migrations for professional reasons are especially lengthy in terms of distance (inter department movements), while family, marriage or accommodation reasons cover shorter distances (within “communes” or counties).

Table 2: Reasons for migration by previous area of living pooled

| | Working | Family | Wedding | Others | |
|--------------------|----------------|---------------|----------------|---------------|--------------|
| From rural | 42.55 | 9.51 | 39.1 | 8.85 | 59367 |
| From urban | 59.93 | 12.48 | 12.02 | 15.57 | 22239 |
| From abroad | 25.53 | 28.14 | 2.86 | 43.48 | 1610 |
| Observation | | | | | 83216 |

Note: The Table represents the reason for migration according to the place of origin of the migrants. The table Shows the percentage by origin and thus should be read in row.

Table 2 shows that migrants previously living in a rural area migrated either for working reasons (42.55% of the time) or for wedding (39.1%). Meanwhile, those living in urban areas migrated mainly for working reasons, reaching almost 60%. More precisely, Bhagat (2010) relates the migrations patterns to urban or rural areas or the other way around. He stresses that the most significant flow is from rural to urban areas, while the reverse stream is the least popular. This peculiar stream is either relative to retired, wealthy people preferring living in the countryside or urban resident, who have to look after the familial properties. Regarding the urban to urban flow, the author highlights that this pattern is essentially concerning the migrants for working reasons (ibid, 2010: 37). Looking at the migration for wedding reason we observe a similar feature to table 1.

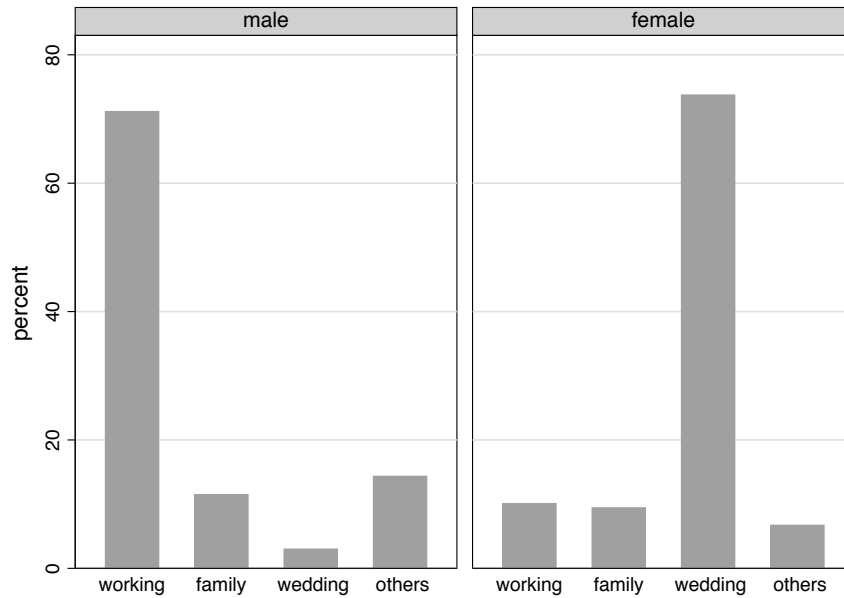
Speaking about working migration Hnatkovska and Lahiri (2013) explain that urban migrant living already in cities earn higher wage than newcomers from rural area.

We can wonder as well how the distribution between women and men is concerning the reasons for migration. We realise this fact on figure 2, migration for wedding reason accounts for 73.75% (very close to Baghat’s estimation of 70%) for women and only 3.01% for men. A reverse situation is observable when we compare migration for professional reason more than 70% of men are concerned against 10.10% of women.

Bhagat (2010) shows very close share of migration reasons relative to men and women, for the Census data in 2001. Despite not finding quantitatively similar result relative to men, qualitatively speaking

we have the same result. The author shows that labour migration seems to be the main drive of men migration (Bhagat, 2010: 39-40).

Figure 2 Migration reasons by gender

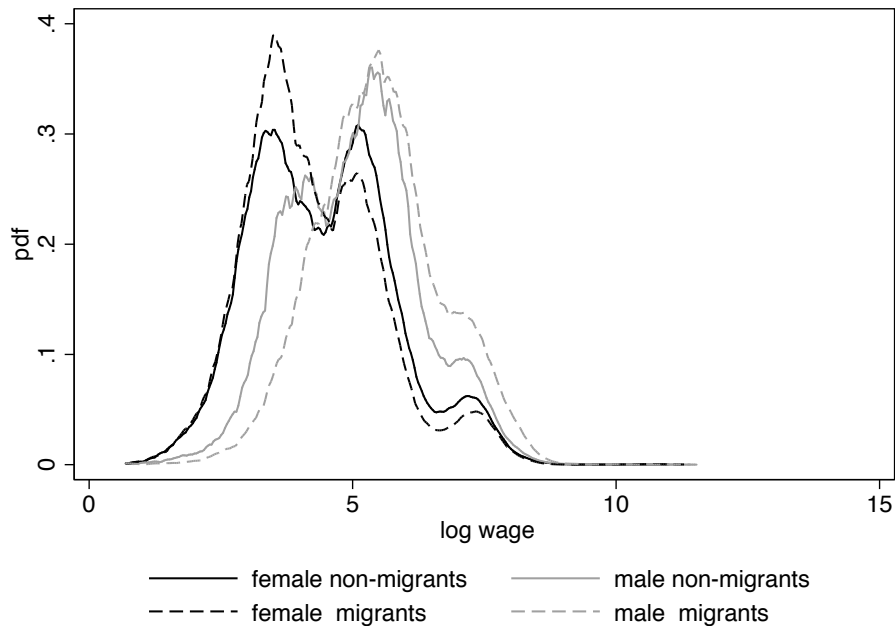


Interestingly, figure 3 shows wage distributions regarding migrants' gender. We notice similar distributions to what we found on figure 1⁸. At a first glance we realise that non-migrants women (solid black) seem not to suffer from a wage penalty of migration in comparison to women who migrated (dashed black). Indeed, the latter is slightly skewed on the left. Concerning men, who migrated (dashed grey), we realise that migrants seems to benefit from a migration's premium relative to the ones who did not migrate (solid grey)⁹. It seems that figure 3 confirms the already mentioned discrimination against women. We will carry more attention to this feature when we will be analysis the Mincer equation in the second part of this section.

⁸ Identically to figure 1 we run a Kolmogorov-Smirnov test allowing us to reject the equality of the distributions. The outcome of these tests are shown in table A.3.2

⁹ Table A.4 gives a clearer view of these distributions relating the descriptive statistics of migration by gender.

Figure 3 Wage distributions for migrants and non-migrants according to the gender



Finally, we wanted to know how the partition of occupations varies between female and male individuals, for both non-migrants and migrants. On figure 4, the left hand side figure depicts the situation for non-migrants. We realise that the main occupation is the elementary occupations for both women and men. On the right hand side the situation is alike for women who migrated but quite different for men. We observe a distribution, which is more balanced among the occupations.

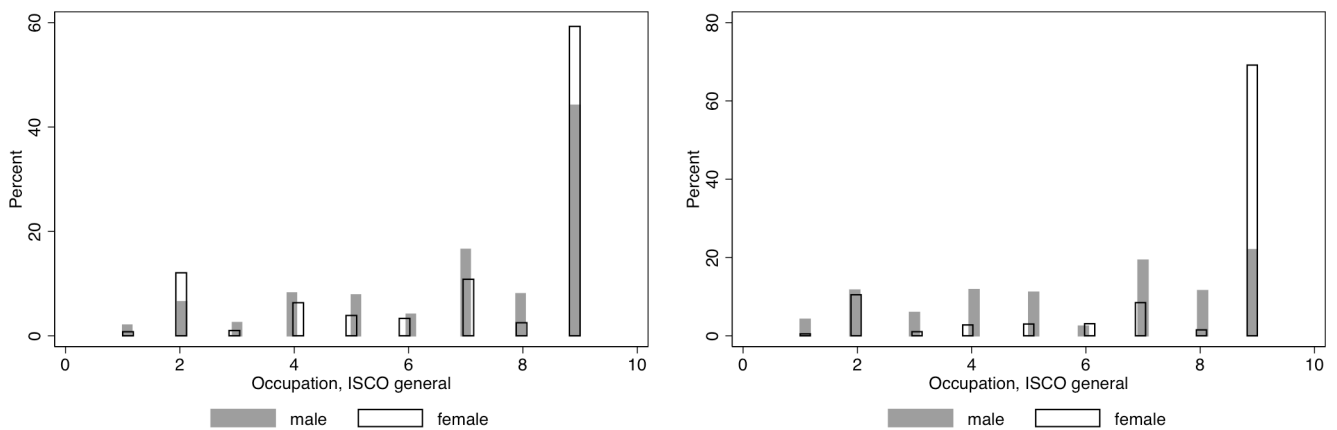
Shanthi (2006) acknowledges that the main migration reason for women is marriage. However, in her article, she underlines that beyond this main reason lie three types of female migration. The first is the migration of *autonomous females*, who fall into one of two categories. One category is that of autonomous female migrants who are highly skilled, while the other includes autonomous women who are not highly skilled. For the former, migration aims either at improving their education or finding suitable employment, with the main motivations being hopes of social advancement and making a better wedding match than was possible in the previous place of living. We suppose that this kind of migrant fits into ISCO categories two, three and four, which correspond to professionals, clerks, and other occupations mentioned figure 4 below. For the latter, semi-literate category of autonomous women, migration goals are usually limited to finding work in manufacturing industries such as the garment industry, electronics assembly lines or the food industry.¹⁰

¹⁰ These categories correspond to 7, 8 and 9 according to ISCO classification, being craft and related trade worker, plant and machine operator and assembler and elementary occupation respectively.

The second type of female migration is labelled *relay migration* and concerns especially elementary occupations (category 9, in figure 4), which remains the principal occupation for women regardless of the status of migration (migrant or non-migrant). In *relay migration*, the eldest daughter is sent to work as housekeeper in order to improve family income until she becomes married, at which point, a younger sister replaces the vacancy left by her elder and so on and so forth. To some extent this second type of migration mirrors Rosenzweig, and Stark (1989) on migration for wedding reason.

Finally, *family migration* is beyond the scope of wedding migration as we categorised this reason for migration under family reason, where women are joining their husband after the wedding expecting to get a job at the destination (Shanthi K., 2006:3).

Figure 4 Occupation by gender for non-migrants and migrants



Left, partition of occupation by gender for non-migrants. Right, partition of occupation by gender for migrants.

4.2 Mincer equation

After having carried a descriptive statistical analysis on the wage premium of the migrants regarding the area of living, the possible reasons for migration and the skills between women and men.

We turn now in the core of our analysis: the comparisons of the wage premia, through Mincer equations.

As we already explained in section 2, the analysis will concentrate first on the simplest form of the Mincer equation, to which we add a set of key control variables (gender, migration status, area of living, occupation, marital status and religion). This equation will be the building block of more complex regressions.

As we explained in the methodology section, we decided to follow Heckman et al. (2003) methodology on imposing a certain number of years of education for the completion of a school degree. In order to gauge the performance of this method, we decided to compare it to the Mincer equation estimation using the years of schooling variable, available only for the year 1983 (equation 2 of table A.5). We realise when we run this equation that the explained models are not so different each other since the coefficients of years in equation (1) and (2) of table A.5 have the same magnitude.¹¹

The second part of our analysis focuses on interactions between migration gender and area of living. We are expecting to get back some features that we observed through figure 1 and 3.

Finally, in the last part we carry the analyses a step further by introducing different motivations for migration and different levels of skill, as explained in section 3. The goal is to determine whether the premia or penalties are specific to precise migrant profiles.

¹¹ This results can be observed in the appendix on table A.5

4.2.1 Mincer equation in its simplest form as a building block

On table 3, we ran four different equations. The first, under column (1), is our building block, which depicts the impact of being a migrant on the wage and fits 36.5% of the data. We developed our regressions by controlling for several variables such as gender, urban area of living, occupation or even religion, which are related under columns (2) to (4). Interestingly, we realised that being a migrant decreases the log wage by 0.086 and is significant at 1% level. Despite being relatively limited, migration is associated with a negative correlation on the wage, *ceteris paribus*.

This result would partially corroborate the wage distributions for migrants and non-migrants that we saw in figure 1, especially for rural migrants. But we do not have the full picture since the migrant's area of living is not taken into account, nor is the migrants' gender wage differential depicted in figure 3. To address this, we ran two other regressions, controlling for gender and urban area (columns (2) and (3) respectively), which explain about 38.8% and 39.6% of the data.

We observe two features, the penalty on migrant living in rural area disappears replaced by a very light premium, about 0.034 and 0.012, significant at 1% and 5% *ceteris paribus*, and a severe penalty hit women with a coefficient about -0.483 and -0.450. It is worth recalling that dummy's coefficient must be interpreted as per cent return.¹²

Although the aforementioned findings support figures 1 and 3, we decided to run a last regression, under column (4), combining all the variables seen so far plus a set of control variables such as occupation and marital status.

By controlling for the set of variables mentioned above we realise that the results stay broadly the same: a light premium for migrant, a heavy penalty for women. Occupations and marital status seem to affect negatively the wage meanwhile religion gives a premium. Indeed as the variable relates the different categories of occupation from the better paid to the less paid (elementary occupations) we observe that the higher the value of this variable the smaller the wage.

Concerning the correlation of an additional year of education or an additional year of experience the coefficient stay roughly the same among the four regressions, about 0.130 for year of education and 0,065 for experience. All the coefficients relative to the last regression are highly significant at 1% level.

¹² As the earning function is on the form of $\ln(y)=\ln(\hat{y}) = \hat{\beta}_0 + \hat{\delta}_i \Leftrightarrow \hat{y} = e^{\hat{\beta}_0 + \hat{\delta}_i}$. Then to get the percent return we evaluate regarding the situation where the dummy takes the value 0 thus we end up with the following expression $\frac{e^{\hat{\beta}_0 + \hat{\delta}_i} - e^{\hat{\beta}_0}}{e^{\hat{\beta}_0}} = e^{\hat{\delta}_i} - 1$.

Table 3 Mincer equations in its simplest form

| | (1) | (2) | (3) | (4) |
|--------------------|----------------------|----------------------|----------------------|----------------------|
| Migrants | -0.086*** (0.005) | 0.034*** (0.005) | 0.012** (0.005) | 0.018*** (0.005) |
| Female | | -0.483*** (0.005) | -0.450*** (0.005) | -0.435*** (0.005) |
| Urban | | | 0.262*** (0.005) | 0.228*** (0.005) |
| Occupations | | | | -0.037*** (0.001) |
| Religion | | | | 0.055*** (0.002) |
| Marital status | | | | -0.077*** (0.005) |
| Years of education | 0.157*** (0.000) | 0.148*** (0.000) | 0.138*** (0.000) | 0.127*** (0.001) |
| Exper | 0.067*** (0.001) | 0.063*** (0.001) | 0.063*** (0.001) | 0.065*** (0.001) |
| Exper2 | -0.001*** (0.000) | -0.001*** (0.000) | -0.001*** (0.000) | -0.001*** (0.000) |
| Constant | 3.289*** (0.009) | 3.454*** (0.009) | 3.379*** (0.009) | 3.540*** (0.017) |
| Observations | 222994 | 222994 | 222994 | 222994 |
| R-squared | 0.365 | 0.388 | 0.396 | 0.401 |
| Year FE | YES | YES | YES | YES |

Note: Standard errors in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% levels respectively. All out regressions are run with year fixed effect.

On table 4, we examine the same set of regressions as in table 3 but we control not only for year fixed effect but also for occupation fixed effect. From now on, we will control for year occupation fixed effects.

The first feature that we can behold is that migration is insignificant under column (1). We also realise that regardless the regressions studied, the penalty for women remains as strong as in the previous table. Despite this fact, the premium for living in an urban area is slightly bigger than in table 3 and remains significant at 1% level.

Notwithstanding, the coefficient of urban variable is indubitably positive regardless the regressions studied. We will try to be even more precise on this aspect, in section 4.2.2, where we will introduce interactions between the area of living, the migration status and the gender.

Table 4 Mincer equations augmented and controlled by year fixed effect

| | (1) | (2) | (3) | (4) |
|--------------------|----------------------|----------------------|----------------------|----------------------|
| Migrants | -0.108 (0.066) | 0.003 (0.024) | -0.006 (0.026) | 0.024*** (0.005) |
| Female | | -0.445*** (0.090) | -0.437*** (0.084) | -0.430*** (0.005) |
| Urban | | | 0.179** (0.070) | 0.251*** (0.005) |
| Religion | | | | 0.058*** (0.002) |
| Marital status | | | | -0.085*** (0.005) |
| Years of education | 0.134*** (0.008) | 0.124*** (0.011) | 0.121*** (0.012) | 0.139*** (0.000) |
| Exper | 0.063*** (0.006) | 0.060*** (0.006) | 0.059*** (0.006) | 0.068*** (0.001) |
| Exper2 | -0.001*** (0.000) | -0.001*** (0.000) | -0.001*** (0.000) | -0.001*** (0.000) |
| Constant | 3.454*** (0.151) | 3.619*** (0.157) | 3.553*** (0.137) | 3.165*** (0.012) |
| Observations | 222994 | 222994 | 222994 | 222994 |
| R-squared | 0.364 | 0.387 | 0.395 | 0.399 |
| Year Occupation FE | YES | YES | YES | YES |

Note: Standard errors in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% levels respectively. All our regressions are run with year-occupation fixed effect.

4.2.2 Interactions between migration, gender and urban area

Table 5 provides a first taste of what is coming next. Indeed, we pushed the analysis a step further from table 4 by looking at the interactions between migrants and urban areas, migrants and gender, and the triple interaction between these variables. The following regressions fit about 40% of the data and are all controlled for year-occupation fixed effects.

Column (1) describes the interaction between migrants and gender. We recover the positive coefficient of being a migrant on the log wage from column (4) of table 4 but we also realise that the association is magnified and still highly significant. The penalty of being women is still there but a bit attenuated and the premium of the urban variable remains strongly significant. Regarding the interaction between migrants and women we observe a feature that we found in figure 3 with the dash black line. Indeed, it seems that women, who migrate suffer from a penalty on the log wage by a coefficient of -0.148, all other things being equal, which increases even more the penalty already present.

On the second column we realise two things, the first is the weakened premium of being a migrant and a tougher penalty of being a women. The second is the positive coefficient of the interaction between

women and urban area, bringing a premium, which foster the premium of living in a city. Moreover all the coefficients of this regression are significant at 1% level.

The regression in column (3) is even more captivating: it shows that being a migrant in a rural area has a negative correlation to income since the coefficient is -0.070 , *ceteris paribus*, significant at 1% level. Moreover, the penalty for being a women and the premium of living in a city are both weakened. However, we realise that the interaction between the migrant and urban area variables is strongly significant and contribute to the premium with a value of 0.173 . Here, we notice two features of figure 1. The first being the penalty on income affecting rural migrants (the black dashed line). The second is about migrants in cities and individuals already living in urban areas, (the solid grey and dashed grey line respectively).

Finally, in column (4) we analyse all the correlations of the interactions that we have seen so far plus we add a triple interaction between migrant, women and urban area.

Interestingly, the coefficient for migrant is negative but not significant. We also recover the same feature that we have studied so far. A severe penalty hit women and a large premium favours those living in an urban area. Despite being significant at 1% level the magnitude of both coefficients is attenuated. As we have seen in column (1), the coefficient of the interaction between migrant and female remains negative. However, being a women and living in a city bring a premium but smaller than in column (2). We also recover the premium of migrant and urban interaction still highly significant and positive.

Concerning the triple interaction about migrants women living in an urban area we realise that the coefficient is positive and significant at 5% level thus bringing a premium on the income of 0.054 .

Through these regressions, we realise that the coefficient relative to migrants is ambiguous. Being positive when we study the interaction with women but negative otherwise. All our regressions clearly show penalties for being a woman, and negative coefficients for the interactions between them and migration remind us figure 3 (columns (1) and (4)). On the contrary, migrants living in cities undoubtedly benefit from a premium, as we have seen in columns (3) and (4).

Despite the fact of having carried analysis on gender urban area and migrant we did not study the kind of migration. Hence, in the next section we are going a step further in the analysis by introducing the reasons for migration and their interactions with the urban variable.

Table 5 Simple interaction and triple interaction between migrants gender and urban

| | (1) | (2) | (3) | (4) |
|---------------------------|----------------------|----------------------|----------------------|----------------------|
| Migrants | 0.067*** (0.006) | 0.033*** (0.005) | -0.070*** (0.007) | -0.003 (0.009) |
| Female | -0.358*** (0.007) | -0.479*** (0.007) | -0.406*** (0.006) | -0.380*** (0.010) |
| Urban | 0.242*** (0.005) | 0.224*** (0.005) | 0.190*** (0.006) | 0.186*** (0.006) |
| Migrants × Female | -0.148*** (0.010) | | | -0.128*** (0.015) |
| Female × Urban | | 0.116*** (0.010) | | 0.045*** (0.015) |
| Migrants × Urban | | | 0.173*** (0.009) | 0.117*** (0.012) |
| Migrants × Female × Urban | | | | 0.054** (0.022) |
| Observations | 222994 | 222994 | 222994 | 222994 |
| R-squared | 0.399 | 0.399 | 0.399 | 0.400 |
| Year- occupation FE | YES | YES | YES | YES |
| Individual controls | YES | YES | YES | YES |

Note: Standard errors in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% levels respectively. All our regressions are run with year-occupation fixed effect. All regressions include individual control variables: religion, marital status, years of education, experience and experience-squared.

4.2.3 Reasons for migration and correlations of occupations

The analysis through the reasons for migration and their interactions

Let's turn to the analysis of the interactions by examining the role of urban areas on the motivations for migration. To answer these questions, we ran four other regressions for simple interactions and one for triple interaction between migration reasons, gender and urban area (respectively columns (1) to (3) and (4)) to determine the premia or the penalties for each reason for migration to urban areas.

In one quick glance at tables 6, we realise that the coefficients concerning the reasons for migration in rural area are all significant at 1% level and negative but migration for working reason. Indeed migration for working reason is the only positive coefficient, regardless the regressions taken into account.

Unsurprisingly, we recover the well-known associations of being a women and living in an urban area. The former being strongly negative whatever the regressions considered and the latter is in every case positive. Concerning the interactions between women and working reason we find ambiguous result. Under column (1), the regression relates a premium about 0.050 significant at 1%, whereas a penalty is observable under column (4). However, this last result is not significant. We also realise that the interaction between women and cities is bringing a premium about 0.095 under column (2) and 0.046, under column (4) when we consider the three-way interaction regression. Suggesting that living in a city allow alleviating the penalty-affecting women. Both coefficients are significant at 1%.

When we focus on the interaction coefficients between the reasons for migration and the urban area (columns (3) and (4)) we appreciate that all the coefficients are positive, regardless of the reason for migration considered. These correlations suggest that urban areas bring a premium on incomes, whatever the motivation taken into consideration. Curiously, the category benefiting the most from this premium in the two regressions considered is the "other reason for migration" variable. We will try to assess this correlation more precisely when we will be studying the three-way interaction including the levels of skill. Almost all the interactions' coefficients are significant at 1% level but for wedding reason, which is at 5%.

Although being in opposition to figure 3, the three-way interaction coefficient seems to corroborate figure 1. Indeed it seems that even if women are suffering from a penalty, the fact of living in a urban area and migrate for working reason bring a premium of 0.150 significant at 1% level, *ceteris paribus*.

Table 6 Interaction between urban area and migrants for working reason

| | (1) | (2) | (3) | (4) |
|--------------------------|----------------------|----------------------|----------------------|----------------------|
| Working | 0.127*** (0.006) | 0.136*** (0.006) | 0.099*** (0.011) | 0.105*** (0.012) |
| Family | -0.067*** (0.014) | -0.078*** (0.011) | -0.146*** (0.018) | -0.141*** (0.024) |
| Wedding | -0.143*** (0.026) | -0.098*** (0.009) | -0.161*** (0.009) | -0.186*** (0.032) |
| Others | -0.074*** (0.012) | -0.087*** (0.011) | -0.173*** (0.016) | -0.164*** (0.019) |
| Female | -0.361*** (0.007) | -0.404*** (0.008) | -0.357*** (0.006) | -0.382*** (0.010) |
| Urban | 0.230*** (0.005) | 0.209*** (0.005) | 0.196*** (0.006) | 0.188*** (0.006) |
| Working × Female | 0.050** (0.020) | | | -0.049 (0.032) |
| Female × Urban | | 0.095*** (0.011) | | 0.046*** (0.015) |
| Working × Urban | | | 0.056*** (0.013) | 0.044*** (0.014) |
| Family × Urban | | | 0.116*** (0.023) | 0.121*** (0.029) |
| Wedding × Urban | | | 0.173*** (0.017) | 0.118** (0.056) |
| Others × Urban | | | 0.161*** (0.022) | 0.165*** (0.025) |
| Working × Female × Urban | | | | 0.150*** (0.041) |
| Observations | 222994 | 222994 | 222994 | 222994 |
| R-squared | 0.401 | 0.401 | 0.401 | 0.402 |
| Year- occupation FE | YES | YES | YES | YES |
| Individual controls | YES | YES | YES | YES |

Note: Standard errors in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% levels respectively. All our regressions are run with year-occupation fixed effect. The reference category is non-migrant for the reason for migration. We represented only the interaction of interest, which are significant. All regressions include individual control variables: religion, marital status, years of education, experience and experience-squared.

These interactions allow us to enhance the presence of a premium for living in an urban area on the log wage. Indeed it seems to go along with the stylized fact that we represent in figure 1, with the grey dashed curve, that migrant living in cities enjoyed better remuneration. Despite not being significant the reason for migration we recover also the penalty-affecting women who migrated.

The analysis through the occupations and their interactions with working migrants

Hitherto, we compared the associations of being a migrant and the coefficient of living in urban area. Then we continue our analysis trying to assess the association of interactions, first on migrant and the area of living and after on the reasons for migration. However, skills have not been formally studied yet. Hence in this section, we will enrich our regression by adding an extra interaction term with different levels of skill.

Let's begin our analysis by looking at the variable without any interactions. As in the previous table, all coefficients associated with reasons for migration but working reason are negative and have about the same magnitude thus we decided to show only this category, on table 7 below. We observe that being a migrant for working reason give a premium for the simple interaction regressions, columns (1) to (3) but shows a negative coefficient of -0.040 significant at 5% level concerning the three-way interaction. Regarding the levels of skill, we unsurprisingly observe that all the coefficients bring a very high premium for highly skilled agents. As a reminder this category described in section 3, contains legislators, professionals charged of producing knowledge and technicians in specific fields.

A premium is also noticeable for intermediate skilled agent (machine operators, skilled agricultural, market sales workers finally clerks) but is not as strong as the previous category. We also retrieve the premium concerning urban discussed all along the tables seen so far. We decide not to represent it on the table but the penalty relative to women is also present.

Curiously, when we analyse interactions between some the most relevant reasons for migration in rural area and skills levels, we realise that the interactions' coefficients between high skilled and working migrant are neither positive nor significant, columns (1) and (4). This result suggests that highly skilled worker suffer from a penalty if they migrate to a rural area. However when we carry our attention on the interaction of this very category with the intermediate skills, we appreciate a positive coefficient of 0.073 and 0.199 significant at 1% level, *ceteris paribus*. Despite the insignificant coefficient relative to the interaction of high skills and working reasons, we have to keep in mind that

the interpretation of interaction is not only about reading the coefficient but is rather a matter of sum of coefficients.¹³

Interestingly, the coefficients between wedding reason for migration in rural area and the levels of skill show reverse results, higher skills migrant seem to benefit from larger premium than intermediate skills one, which coefficients of the interaction are negative.¹⁴

While coefficients concerning the interactions between the “other reasons for migration” variable and the levels of skill are positives in any case, all the correlations discussed in this paragraph are significant at least at 5% level.

Regarding the interactions between reasons for migration and urban area we notice very close results to those found in table 6. All the coefficients are significant at 1% level, regardless the reasons considered in column (3). However the correlations are not so sharp concerning the three-way interaction regression, column (4). Indeed only working and family reasons are positive and significant.

The analysis of the interaction between the levels of skill and the urban area gives unexpected results, we would have expected the interactions coefficients to foster the premium on the income but we rather observe the opposite with highly negative coefficients in column (3) and even more severe concerning column (4)¹⁵. A possible explanation of these negatives coefficients could be the fierce competition among worker in cities that would decrease the income but this phenomenon would rather affect low skilled worker rather than highly skilled ones.

Finally, we carried out the analysis on three-way interaction. Increasing the number of interaction complicates the interpretations not only because of the variables themselves but also all the simple interactions between them must be taken into account to assess the impact correctly. Hence, we will only comment on the magnitude and the significance of these triple interactions. We would have expected the coefficient concerning the interaction between working reason, urban area and skills to be positive but we get instead a coefficient about -0.180 for the relation relative to intermediate skill, significant at 1%, *ceteris paribus*, while the interaction with high skills is not significant.

Concerning the correlations relative to wedding interaction and other reasons interaction. The former shows expected positive coefficients of 0.264 and 0.381 for both high and intermediate skill respectively, while the latter shows a coefficient of 0.253 regarding the intermediate skill. All these coefficients are significant at 1% level, all other things being equal

¹³. Indeed to catch the correct association we should sum the coefficients of the working reason for migration and high skill variables plus the interaction between the two. Hence, we realise that the interaction does not bring any additional value to the addition since is not significant. Concerning the regression under column (1) the relation will bring a premium of 0.212 (0.077+0.135)

¹⁴ For column (1) the value of the premium would be 0.478 (-0.177+0.320+0.275) for the relation regarding the interaction with high skills and -0.122 (-0.177+0.135-0.100) for intermediate skills.

¹⁵ The premium for the relation between high skills and urban is 0.511 and 0.496, respectively for column (3) and (4), while those for intermediates are 0.350 and 316 respectively for column (3) and (4).

Table 7 Interactions between migrant for working reason and occupations

| | (1) | (2) | (3) | (4) |
|---------------------------------------|----------------------|----------------------|----------------------|----------------------|
| Working | 0.077*** (0.012) | 0.055*** (0.011) | 0.107*** (0.006) | -0.040** (0.017) |
| Wedding | -0.117*** (0.009) | -0.154*** (0.009) | -0.106*** (0.008) | -0.115*** (0.010) |
| Others | -0.223*** (0.018) | -0.189*** (0.016) | -0.100*** (0.011) | -0.244*** (0.021) |
| High skills | 0.320*** (0.010) | 0.331*** (0.009) | 0.418*** (0.013) | 0.471*** (0.016) |
| Intermediate skills | 0.135*** (0.006) | 0.152*** (0.006) | 0.179*** (0.007) | 0.192*** (0.009) |
| Urban | 0.194*** (0.005) | 0.154*** (0.006) | 0.258*** (0.008) | 0.238*** (0.010) |
| Working × High skills | -0.010 (0.018) | | | -0.048 (0.031) |
| Working × Intermediate skills | 0.073*** (0.014) | | | 0.199*** (0.025) |
| Wedding × High skills | 0.275*** (0.026) | | | 0.085** (0.040) |
| Wedding × Intermediate skills | -0.110*** (0.018) | | | -0.279*** (0.022) |
| Others × High skills | 0.192*** (0.031) | | | -0.020 (0.053) |
| Others × Intermediate skills | 0.191*** (0.024) | | | 0.132*** (0.034) |
| Working × Urban | | 0.087*** (0.013) | | 0.185*** (0.024) |
| Family × Urban | | 0.125*** (0.023) | | 0.110*** (0.038) |
| Wedding × Urban | | 0.176*** (0.017) | | 0.005 (0.024) |
| Others × Urban | | 0.169*** (0.021) | | 0.060 (0.041) |
| High skills × Urban | | | -0.165*** (0.015) | -0.240*** (0.020) |
| Intermediate skills × Urban | | | -0.087*** (0.010) | -0.114*** (0.013) |
| Working × High skills × Urban | | | | 0.029 (0.038) |
| Working × Intermediate skills × Urban | | | | -0.184*** (0.032) |
| Wedding × High skills × Urban | | | | 0.264*** (0.056) |
| Wedding × Intermediate skills × Urban | | | | 0.381*** (0.040) |
| Others × High skills × Urban | | | | 0.253*** (0.070) |
| Observations | 222734 | 222734 | 222734 | 222734 |
| R-squared | 0.406 | 0.405 | 0.405 | 0.407 |
| Year- occupation FE | YES | YES | YES | YES |
| Individual controls | YES | YES | YES | YES |

Note: Standard errors in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% levels respectively. All our regressions are run with year-occupation fixed effect. The reference category is non-migrant for the reason for migration. And low skills for the levels of skill. We represented only the interaction of interest thus we take out experience squared gender religion and marital status being similar to what we had in the previous table. All regressions include individual control variables: religion, marital status, years of education, experience and experience-squared.

5. Concluding remarks and discussion

In this section we will review our results; then we will present some of the study's limitations and discuss some aspects of urban economics related to migrations through the lens of recent literature. A general remark that we want to emphasize is that all the results being presented do not aim to identify causality between the explanatory variable and the income of migrants, but rather to give insight on the relation among these variables.

Our first finding concerns the premium of urban-benefiting male migrants, regardless of their reasons for migration. We found by employing two different methods. On the one hand descriptive statistics outlined what migration looks like for men, as shown in figures 1, 3 and 4. Indeed, figure 1 seems to favour migrating toward urban areas, while figure 3 seems to support men's migration. On the other hand, the applications of Mincer equations allowed us to quantify different premia. Table 6 provides early insight on the premium the variable is analysed without considering any interactions. At first glance, working reasons appear to be the sole reasons offering a premium when the migration occurs towards rural areas.

However, a study of the interactions between the reasons for migration urban areas yields the expected results, since the coefficients are not only positive but also highly significant. Among all the motivations for migration, "other reasons" are the ones which present the highest positive coefficient. Even if these relations concerning simple interactions are covered in table 7, three-way interaction shows the evidence is not as clear as in the previous table suggests, since both wedding and other reasons for migration show positive coefficients.

Having assessed the evidence on male migration, we moved on to its corollary, the examination of the correlations of urban variables through the same methodology. Hence, our second finding sheds light on women's migration. Despite suffering from their gender status, migrations toward cities seem to alleviate the penalty on women's income. Figure 3 shows that female migrants suffered from a penalty compared to their counterparts who did not migrate. However, this figure should be completed by other information, since it does not give any context about migration to urban areas. What can be said is that women are worse off when they migrate. (Table 5 shows the penalty which affects women.) Furthermore, this negative link is exacerbated when examining the interaction between the status of migration and women in columns (1) and (4). In spite of this, three-way interactions coefficients between migration, women and urban area seem to show some evidence of a premium alleviating this penalty. A similar observation can be made about column (4) of table 6, which examines the different reasons for migration.

The last result that we expected to find was an important premium on income brought by migrant skill level, especially in cities. This finding is assessed in the very last table of section 4, through regressions concerning migration and levels of skill, using low skills as a reference category (according to the distribution of occupation that we detailed in section 3). Unsurprisingly, higher skill level variables are linked to a very large premium regardless of the regression.

However, on the simple interactions between migrants' motivation and skills level, the results are mixed. The striking findings that we have to stress from these simple interactions are the ones concerning wedding and other reasons for migration because they present the expected correlations.

Regarding the triple interactions among the reasons for migration, urban areas and skills, we expected high and positive coefficients on the income for working reasons, but we observe the opposite, since heavy penalties deeply affect migrants for professional reasons. (column (3) and (4) of table 7.) Surprisingly, we also recover a large premium for wedding and other migrations on the triple interaction. The former finds explanations in literature. Indeed, Shanti (2006) explains that upper class women migrated to urban areas to improve their education or to find suitable employment, hence, raising the likelihood to have a better marriage and expect social advancement. The premium that we capture might be the reflection of this explanation. One interpretation of the positive coefficients for the "other reason" category is that since it is mainly composed of migrants who move for education reasons and that this type of migrant is likely to move to more developed states, benefit from higher quality education and infrastructure, they benefit from higher future income (Chandrasekhar and Sharma (2014)).

An alternative explanation could be one given by Khan (2016). The article assesses that migrants with technical educations earn high incomes than their counterparts without technical education and this might be due to a better ability of migrants to acquire skills at destination.

Through all these analyses carried on table 7, migrations reason and urban variables seem to be the driver than the premium are brought rather by the migration and urban variable than by the levels of skill concerning working reason for the three-way interaction. Indeed, observations relative to simple interactions corroborate figure 1 where the premium for migrant living in urban area is triggered by the urban variable.

Decomposing the migration variable by reasons for migration allows us to have a clearer view at least for simple interaction between migrant and levels of skill. Concerning the triple interaction, the correlations are blurrier because of the penalty affecting working migrant while other migrants seem to enjoy a premium.

5.1 Limitations

An important limitation of our study is a bias toward endogeneity, even though a set of independent variables was used to explain the income differences between migrants and non-migrants as an attempt to correct for this bias. However reverse causality can also be a source of bias: are decisions to migrate made because of the prospect of higher income, or does higher income allow migrants to decide whether to migrate or not? The use of the instrumental variables method might also be a solution as long as the instruments are uncorrelated with other latent variables and succeed in breaking the endogeneity without generating other problems. (Card, D., 1999: 1819-1822).

Selection effect impacts our results as well, as by focussing only on remunerated and active agents, we exclude from our sample every agent working but not being paid along with individuals who get paid without working. Furthermore, migrants are a very specific group of people who choose to migrate according to observable and non-observable characteristics. As Khan (2016) explains, wage regressions usually present selection bias, since comparing the coefficients of migrants and non-migrants workers makes clear that migration decisions are not a random process

Our datasets for the years 1983, 1987 and 1999 were not able to distinguish between permanent and temporary migration. Since it has been established that temporary migration is seven times more common than permanent migration (Keshri and Bhagat (2013)), our inability to segregate the two also constitutes a limitation to our study. Important progress has been made on more recent surveys such as the one carried out by the National Family Health Surveys in 2005-2006 (NFHS-3) including a specific question on time spans for continuously living in the same place of residence (Bhagat (2010)). In the future, such progress will allow more precise analysis of the reasons for migration.

5.2 Related field and further research

Rural to urban migration captures most of our attention because it embodies important transformations of the structure of developing economies, as they pass from primarily agricultural activities to being dominated by industry and services. When any development process occurs, congestion effects are expected, triggering pressure on social services in the cities and downward pressure in wages on the labour market (Lucas (2015)). Contrary to China, where a permit system (hukou) restricts residence in urban areas (Hering and Poncet (2010)), Indian citizens enjoy freedom of movement.

Out of the scope of India, but more generally in urban economics, research has been carried out to assess the sorting of migrants toward cities; this not only for the most skilled, but also the least skilled individuals, which are to a certain extent observably complementary (Eeckhout et al. 2014 and Behrens et al. 2014). De la Roca (2011) examines the effects of cities on migration and explains that migrants might return to their previous destination if they miscalculated the benefit that they could gain. By the same token, De la Roca et al. (2014) explains that migration is also a sorting process

according to the ability of migrants, who migrate to bigger cities if they believe that they would have greater benefit than in smaller cities. Despite being essentially focused on urban to urban migration the aforementioned research might also be carried out in India in the near future, provided that the upgrade of the databases continues.

5.3 Conclusion

This paper presents new evidence on the relationships between the reasons for migration toward urban areas and income in India, casting light on the reasons for migration. Indeed cities confer a premium to migrants, which not only differs a lot depending on the reason for migration but also allows a particular alleviation of discrimination against women.

A two-step approach was used to determine the benefits or penalties of migration. The study turned first to descriptive analyses based on the distributions of wages between migrants and non-migrants and between women and men. Then the analysis was complemented by Mincer equations evaluating the impact of migration on income. These equations were then progressively supplemented by additional variables and interactions.

The main findings are in line with literature results. Urban areas benefit unambiguously to male migrants regardless of the reasons taken into account. However, when the regressions become more complex the correlations are not so clear especially concerning individuals migrating to urban areas for work reasons. Nonetheless, the premium seems to be reinforced for wedding and other reasons for migration (mainly education). Despite being counter-intuitive, these results find explanations in the literature. Furthermore, we found heavy penalties affecting female migrants, but migration toward cities seem to lessen them.

A novelty of this study is the attempt to link two threads of the literature: Khan's (2016) article examining the wage differential between migrants and non-migrants and the more qualitative Bhagat (2010) article focused on features of the reasons for migration in India.

References

- Abbas, R. and Varma, D. (2014). Internal Labor Migration in India Raises Integration Challenges for Migrants. *Migration Information Source*.
- Banerjee, B. (1981). Rural-urban migration and family ties: an analysis of family considerations in migration behaviour in India. *Oxford bulletin of economics and statistics*, 43(4), 321-355.
- Behrens K., Duranton G. and Robert-Nicoud F. (2014). Productive Cities: Sorting, Selection, and Agglomeration. *Journal of Political Economy*, 122 (3), 507-553.
- Bhagat, R. B. (2010). Internal migration in India: are the underprivileged migrating more. *Asia-Pacific Population Journal*, 25(1), 27-45.
- Bertoli S., Fernandez-Huertas Moraga and Ortega F. (2013). Crossing the Border: Self-selection, Earnings and Individual Migration Decisions. *Journal of Development Economics*, 101, 75-91.
- British Council (2014). Indian School Education System An Overview. Online: https://www.britishcouncil.in/sites/default/files/indian_school_education_system_an_overview_1.pdf [last seen 26.06.2017].
- Cahuc, P. and Zylberberg, A. (2004). *Labor Economics*. Cambridge, MA: MIT Press.
- Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics*, 3, 1801-1863.
- Census of India 1981 (1988). Online : http://www.censusindia.gov.in/data/Census_1981/Publication/India/49962_1981_POR.pdf
- Chandrasekhar, S. and Sharma, A. (2014). Internal migration for education and employment among youth in India. *Indira Gandhi Institute of Development Research, Mumbai January*.
- Czaika, M. (2012). Internal versus international migration and the role of multiple deprivation: some evidence from India. *Asian Population Studies*, 8(2), 125-149.
- De la Roca, J. (2011). Selection in initial and return migration: Evidence from moves across Spanish cities. *Processed, cemfi and imdea Social Sciences*.
- De la Roca, J., Ottaviano, G. and Puga, D. (2014). City of dreams. CEP Discussion Papers, CEPDP1305. Centre for Economic Performance, London School of Economics and Political Science, London, UK.
- Eeckhout, J., Pinheiro, R. and Schmidheiny, K. (2014). Spatial Sorting. *Journal of Political Economy*, 122 (3), 554-620.
- Fafchamps, M. and Shilpi, F. (2013). Determinants of the Choice of Migration Destination. *Oxford Bulletin of Economics and Statistics*, 75 (3), 388-409.
- Gobillon L. (2001). Emploi, logement et mobilité résidentielle. *Economie et statistique*, 349(1), 77-98.
- Greenwood, M. J. and Hunt, G. L. (2003). The early history of migration research. *International Regional Science Review*, 26(1), 3-37.

- Haan, A. D. (1997). Rural-Urban Migration and Poverty: The Case of India. *IDS Bulletin*, 28(2), 35-47.
- Harmon, C., Oosterbeek, H. and Walker, I. (2003). The returns to education: Microeconomics. *Journal of economic surveys*, 17(2), 115-156.
- Head K. and Mayer T. (2004). Market Potential and The Location of Japanese Investment in the European Union. *The Review of Economics and Statistics*, 86 (4), 959-972.
- Heckman J. J., Lochner, L. J., and Todd, P. E., (2003). Fifty Years Of Mincer Earnings Regressions, *NBER Working Paper 9732*, 1-73.
- Hering, L., and Poncet, S. (2010). Market access and individual wages: Evidence from China. *The Review of Economics and Statistics*, 92(1), 145-159.
- Hnatkovska, V. and Lahiri, A. (2013). Structural transformation and the rural-urban divide. *University of British Columbia, typescript*.
- Humphreys, J. (2013). An alternative to the Mincer model of education. In *42nd Australian Conference of Economists, Perth, 10th July*.
- ISCO. (2010). Online <http://www.ilo.org/public/english/bureau/stat/isco/isco88/publ4.htm> [last seen 03.01.2017]
- Keshri, K. and Bhagat, R. B. (2013). Socioeconomic determinants of temporary labour migration in India: A regional analysis. *Asian Population Studies*, 9(2), 175-195.
- Khan, M. I. (2016). Migrant and non-migrant wage differentials: a quintile decomposition analysis for India. *The Indian Journal of Labour Economics*, 59, 245-273.
- Lucas, R. E. (2014). Internal migration in developing economies: An overview. *Paper produced for KNOMAD's Thematic Working Group (TWG) on Internal Migration and Urbanization. KNOMAD*.
- Lusome, R. and Bhagat, R. B. (2006). Trends and patterns of internal migration in India, 1971–2001. In *Annual conference of Indian Association for the Study of Population (IASP) during* (Vol. 7, p. 9).
- Mahapatro, S. (2014). Contemporary Patterns and Issues of Internal Migration in India: Evidence from NSSO. In *KNOWMAD conference on Internal Migration and Urbanization, Dhaka, May 1st*.
- Mayneris, F., Poncet, S., & Zhang, T. (2016). *Improving or Disappearing: Firm-Level Adjustments to Minimum Wages in China* (No. 2016027). Université catholique de Louvain, Institut de Recherches Economiques et Sociales (IRES).
- Mincer, J. (1974). *Schooling, Experience and Earnings*. New York: National Bureau of Economic Research.
- Mincer, J. (1978). Family migration decisions. *Journal of political Economy*, 86(5), 749-773.
- Minnesota Population Center. (2017). *Integrated Public Use Microdata Series, International: Version 6.5* [dataset]. Minneapolis: University of Minnesota, 2017. Online: <http://doi.org/10.18128/D020.V6.4>. <https://international.ipums.org/international/index.shtml> [last seen 10.07.17]
- Moretti, E. (2012). *The new geography of jobs*. Houghton Mifflin Harcourt.
- National Sample Survey Organisation. (1988). Online:

http://www.mospi.gov.in/sites/default/files/publication_reports/nss_report_347.pdf[last seen 14.07.2017]

Rao, N. and Rana, K. (1997). Women's labour and migration: the case of the Santhals. *Economic and Political Weekly*, 3187-3189.

Rosenzweig, M.R., and Stark, O. (1989). Consumption Smoothing, Migration, and Marriage: Evidence from Rural India, *Journal of Political Economy*, 97(4), 905-926.

Shanthi, K. (2006). *Female labour migration in India: Insights from NSSO data* (Vol. 4). Madras School of Economics.

Srivastava, R., and Sasikumar, S. K. (2003). An overview of migration in India, its impacts and key issues. In *Regional Conference on Migration, Development and Pro-Poor Policy Choices in Asia*

Stark, O. (1984). Rural-to-urban migration in LDCs: a relative deprivation approach. *Economic Development and Cultural Change*, 32(3), 475-486.

Stark, O. and Yitzhaki, S. (1988). Labour migration as a response to relative deprivation. *Journal of Population Economics*, 1(1), 57-70.

Thorat, S., and Attewell, P. (2007). The Legacy of Social Exclusion: A Correspondence Study of Job Discrimination in India. *Economic and Political Weekly*, 42(41), 4141-4145.

United Nations Population Division. (1970). *Methods of Measuring Internal Migration. Manuals on Methods of Estimating Population. Manual VI*, UN, New York.

Appendix

Table A.1 Shares of reason for migration regarding the non remunerated agents

| | Non-migrant | Working | Family | Wedding | Others | |
|--------------|-------------|---------|--------|---------|--------|--------|
| Male | 89.75 | 4.57 | 2.07 | 0.68 | 2.93 | 282467 |
| Female | 43.67 | 0.93 | 3.12 | 50.29 | 1.98 | 95326 |
| Observations | | | | | | 377793 |

Note: the table shows the percentage by row of the non-remunerated active agent by gender.

Table A.2 Wage distribution statistics by area of living

| | mean | median | skewness | kurtosis | obs. |
|--------------------|--------|--------|----------|----------|-------|
| Rural non-migrants | 4.6098 | 4.6151 | 0.2195 | 2.7062 | 74203 |
| Rural migrants | 4.3789 | 4.2627 | 0.4551 | 3.3184 | 38081 |
| Urban non-migrants | 5.3180 | 5.3033 | 0.0046 | 3.1449 | 65315 |
| Urban migrants | 5.5114 | 5.4848 | -0.0084 | 3.0131 | 45395 |

Table A.3.1 Two-sample Kolmogorov-Smirnov test for equality of distribution between rural and urban area

| Rural | D | P-value | Urban | D | P-value |
|---------------|---------|---------|---------------|---------|---------|
| Non-migrants: | 0.0037 | 0.499 | Non-migrants: | 0.0676 | 0 |
| Migrants: | -0.1032 | 0.000 | Migrants: | -0.0001 | 0.999 |
| Combined K-S: | 0.1032 | 0.000 | Combined K-S: | 0.0676 | 0 |

Note: the table shows the outcome of the Kolmogorov-Smirnov test. The result of interest is the third line we realise that in any case we reject the null supposing that the distribution are equal

Table A.3.2 Two-sample Kolmogorov-Smirnov test for equality of distribution between genders

| Male | D | P-value | Female | D | P-value |
|---------------|---------|---------|---------------|---------|---------|
| Non-migrants: | 0.145 | 0 | Non-migrants: | 0.0006 | 0.989 |
| Migrants: | -0.0002 | 0.998 | Migrants: | -0.0837 | 0 |
| Combined K-S: | 0.145 | 0 | Combined K-S: | 0.0837 | 0 |

Note: the table shows the outcome of the Kolmogorov-Smirnov test. The result of interest is the third line we realise that in any case we reject the null supposing that the distributions are equal

Table A.4 Wage distribution statistics by gender

| | mean | median | skewness | kurtosis | obs. |
|--------------------|----------|----------|-----------|----------|--------|
| Women non-migrants | 4.444904 | 4.382027 | 0.3587239 | 2.835854 | 22885 |
| Women migrants | 4.262368 | 4.077538 | 0.6174455 | 3.49315 | 33159 |
| Men non-migrants | 5.038748 | 5.081404 | 0.0970934 | 2.849246 | 116633 |
| Men migrants | 5.477469 | 5.442418 | 0.0947756 | 3.043624 | 50317 |

Table A.5 Years of schooling versus year proxy

| | (1) 1983 | (2) 1983 |
|--|----------------------|----------------------|
| Year of education as the highest grade completed | 0.137*** (0.001) | |
| exper | 0.056*** (0.001) | |
| exper2 | -0.001*** (0.000) | |
| Years of schooling | | 0.134*** (0.001) |
| experbis | | 0.059*** (0.001) |
| experbis2 | | -0.001*** (0.000) |
| Constant | 2.816*** (0.011) | 2.751*** (0.010) |
| Observations | 87061 | 87061 |
| R-squared | 0.415 | 0.426 |

Note: Standard errors in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% levels respectively. Both regressions are run with the data concerning 1983