Where Does the Stigma Lie? : Exploring the Roles of Gender, Religion and Caste in the Indian Labor Market

Suneha Seetahul

▶ To cite this version:

Suneha Seetahul. Where Does the Stigma Lie? : Exploring the Roles of Gender, Religion and Caste in the Indian Labor Market. Economics and Finance. Université de Bordeaux, 2018. English. NNT : 2018BORD0337 . tel-02076981

HAL Id: tel-02076981
https://theses.hal.science/tel-02076981

Submitted on 22 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE
L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE ENTREPRISE, ÉCONOMIE, SOCIÉTÉ (N°42)
SPÉCIALITÉ SCIENCES ÉCONOMIQUES

Par Suneha SEETAHUL

Where Does the Stigma Lie?

Exploring the Roles of Gender, Religion and Caste in the Indian Labor Market.

Sous la direction de : François COMBARNOUS

Soutenue le 30 novembre 2018

Membres du jury :

Mme HARRISS-WHITE Barbara

Emeritus Professor, University of Oxford, ***Présidente du jury***

M. LANJOUW Peter F.

Professor, Vrije Universiteit Amsterdam, ***Rapporteur***

M. NORDMAN Christophe J.

Chargé de recherche, Institut de Recherche pour le Développement, DIAL et Institut Français de Pondichéry, ***Rapporteur***

M. BERNARD Tanguy

Professeur des Universités, GREThA, Université de Bordeaux, ***Examinateur***

Mme. RANI Uma

Senior Development Economist, International Labour Organization, ***Examinatrice***

M. COMBARNOUS François

Maître de conférences HDR, GREThA, Université de Bordeaux, ***Directeur de thèse***

Titre : **Une analyse socioéconomique du genre, de la religion et de la caste sur le marché du travail indien**

Résumé : Cette thèse propose de détecter les mécanismes par lesquels la stratification d'une société se reflète sur le marché du travail. Nous étudions pour cela le cas de l'Inde, où les disparités liées au genre, à la religion et à la caste persistent malgré des changements structurels considérables. Un premier chapitre traite des liens entre l'exclusion du marché du travail et les disparités inter-groupes. Suite à une estimation de la probabilité de non-participation sur le marché du travail, l'analyse des conséquences de l'isolement forcé des femmes sur l'éducation des enfants permet d'observer dans quelle mesure le travail est un vecteur de réduction des inégalités genrées d'éducation. Un second chapitre s'intéresse aux trajectoires de mobilités occupationnelles et de revenu entre 2005 et 2011-12. Une description détaillée de ces trajectoires et l'estimation de ces déterminants suggèrent l'absence d'un phénomène de rattrapage des groupes désavantagés sur le moyen-terme. Un troisième chapitre aborde la question de la segmentation du marché du travail dans un contexte de prédominance de l'économie informelle. Les résultats montrent l'existence d'une ségrégation occupationnelle en fonction du genre. Le quatrième chapitre propose une comparaison paramétrique et non-paramétrique des décompositions de salaire. Les écarts de salaires ne sont pas liés à une discrimination salariale pure mais plutôt à un processus de sélection et de ségrégation occupationnelle dans le cas du genre. En ce qui concerne les groupes socio-religieux, la combinaison des écarts en termes d'éducation, de népotisme et de discrimination potentielle explique les fortes disparités salariales.

Mots clés : Genre, Caste, Marché du travail, Discrimination, Segmentation, Inde

Title: **Where Does the Stigma Lie? Exploring the Roles of Gender, Religion and Caste in the Indian Labor Market.**

Abstract: This thesis aims to analyze how the stratified nature of a society translates into horizontal inequalities in the labor market. We analyze the case of urban India, where disparities among gender, religion and caste groups persist despite the country's significant structural change. The first chapter analyzes the links between labor market exclusion and group disadvantage. After estimating the likelihood of non-participation in the labor market, we address the specific case of secluded labor by detecting its impact on children's education. We suggest that female labor market participation is not likely to lower the educational gap for future generations. The second chapter compares the paths of labor market mobility between 2005 and 2011-12 among gender and socio-religious groups. A detailed analysis of occupational and earnings mobility, followed by the estimation of their determinants, suggest that the group-specific mobility patterns may not reflect a process of "*catching-up*." The third chapter proposes an analysis of labor market segmentation in the context of a predominantly informal labor market, showing that the household business sector is relatively homogenous and that the salaried sector is segmented along gender lines. A fourth chapter highlights the issue of potential discrimination by comparing parametric and semi-parametric wage gap decompositions, both suggesting that wage gaps are mostly due to selection and segregation effects in the case of gender. In the case of socio-religious groups, a combination of endowment differentials, nepotism and potential discrimination leads to substantial wage differentials.

Keywords: Gender, Caste, Labor Market, Discrimination, Segmentation, India

GREThA UMR CNRS 5113 – Université de Bordeaux

Avenue Léon Duguit, 33 608 Pessac Cedex

Acknowledgements

I would like express sincere thanks to those who helped this project become a thesis and who contributed to making this experience successful.

I would first like to thank the esteemed members of the jury, Mrs. Barbara Harriss-White, M. Lanjouw Peter, M. Nordman Christophe J., Mr. Bernard Tanguy and Mrs. Rani Uma for accepting to be a part of my defense committee and reviewing my thesis.

I address my sincere gratitude to François Combarnous, my Ph.D. advisor, for his trust, support and precious comments. Your course of *Economie du Développement* during my third-year Bachelor's degree, motivated me to pursue in this field. Your passion for teaching and development is inspiring.

I would like to thank the French Institute of Pondichéry for their welcome and support. I address my kind regards to Kamala Marius and Venkatasubramnian for giving me the opportunity to collaborate with them.

I also thank all of the members of the GREThA. In particular, I am grateful to Marc-Alexandre Sénégas for his support in organizing the AJEI 2018 workshop. I also thank the GREThA development team, Matthieu, Claire, Eric R., Eric B., André, Tanguy, as former teachers, and current colleagues for their availability and for trusting me with the organization of the Séminaire d'Economie du Développement in 2016.

I thank Riana, Pierre, Nico B., Nico Y., Pauline, Viola, Valentina, Christian, Keshav, Dimi and Yashou for helping me with to proofread this thesis.

This journey would have been very different without my colleagues and friends from the GREThA. The tremendous regularity of the pause café is going to be missed! Pierre, your good mood and bad jokes made the gloomy *travée D* a fun place. I wish you the best, now that stardom awaits you. Youlouch and Pauline, I thank you for your constant support and friendship. I wish to address a special thanks to Riana for being my acolyte in the last months. Your support and friendship mean a lot to me. Thanks, you Sébastien and Thibaud for your advice. I wish the best to François and Dan for the few days to come. Good luck to Louis, Badr, Jeanne, Lucie, Erwan, Elodie, Coralie, Fadoua, Adil, Jérémy, to 80% Marxist-feminist Lucille (I wonder where the rest went), to Marjojo for showing us that you can and you should follow your dreams, and to all of the PhD students of the GREThA.

I thank the Cheverus bunch for all the fun times. No matter what we say, we are creatures of habit! My dearest Viovio, thank you for everything, your crossing-over from the Larefi to the GREThA was the best thing that happened two years ago. Bedu, thanks for the continuous PhD mentoring, you're such a pro at this! Junior, I won't wish you good luck, as you may have already finished your dissertation since I last saw you.

Thanks to my friends for putting up with my academic whining in the last months (or years)! Specially to Cécilia, Kevin, Soso, Diane, Juliette, Sarah, Enek, Cece, Lolo, osbivians oldies and newbies.

Mathilde, Ninie, Cey and Alex, writing this thesis was bound to make me remember our years of discovering and loving India.

Yashvin, I salute your perseverance and hard work. You deserve the best and I have no doubt that you will achieve it. Maman, thank you for your dedication, sacrifice and generosity. You both mean the world to me and without your support I would definitely not be here.

When I arrived in Bordeaux in September 2008, I told my dad that I would probably stay here for a decade and end up doing a PhD as a joke. It turns out I was not wrong. I wish he were here to read this thesis. I dedicate this work to you, Papa.

Last but not the least, my dearest Dimi, you finally did get your paragraph in the acknowledgements, although you deserve much more. You have my most sincere gratitude for your patience and constant support. We have been through so many life steps in these last years and I can't wait to embark on a new adventure with you.

Table of contents

General Introduction

"It is through work that women have been able, to a large extent, to close the gap separating her from the male, work alone can guarantee her concrete freedom."

Simone de Beauvoir, Le deuxième sexe (1949)[1]

"On the 26th of January 1950, we are going to enter into a life of contradictions. In politics we will have equality and in social and economic life we will have inequality.
In politics we will be recognizing the principle of one man one vote and one vote one value.
In our social and economic life, we shall, by reason of our social and economic structure, continue to deny the principle of one man one value."

Bhimrao Ramji Ambedkar, Last speech before the Constituent Assembly in 1949[2]

Equality being a key component of economic development, public policy needs to ensure that all socioeconomic groups are offered the same opportunities and are equally able to seize them. The persistence of socioeconomic inequality across groups characterized by a common gender, religion or caste group, combined to their apparent specificity regarding work-related mechanisms, crystallize economic disparity on the Indian labor market. Both quotes at the beginning of this introduction date from 1949, and yet, they remain of prime pertinence in contemporary India. By shedding new light on the extent and the nature of horizontal inequalities based on gender, religion and caste in the Indian labor market, this thesis explores whether these dimensions constitute primary structuring factors of the labor market.

[1] Author's translation

[2] Retrieved from Drèze and Sen (2013). Bhimrao Ramji Ambedkar was an Indian politician and jurist. He led the committee that drafted the Constitution.

1. A general characterization of the Indian economy

The Indian economy has attracted much attention in the last decades. Diverging labels such as "*emerging country,*" "*BRIC*" or "*Lower-Middle Income Country*" are used to qualify the country. These terms testify to the economy's specificity as a transition economy but also show a discrepancy between the political will to be a key player in the globalized economy and the socioeconomic reality of insufficient per capita income. The long-term trends show optimistic perspectives for economic development (The World Bank 2018), with an average annual growth rate of 6.6% between 2011 and 2017 (Woetzel, Madgavkar, and Gupta 2017). However, the poverty rate is still high (21.6% in 2016[3]), with 21.2% of working poor individuals in 2011 (Asian Development Bank 2018). The schizophrenic image of the country is often emphasized, opposing an India that benefits from growth to an India that stays in situations of poverty and vulnerability (Harriss-White 2003; Boillot 2016). It makes no doubt that the country has undergone significant structural change since the 1990s. The shift in the structure of the economy, from the agricultural sector to industry and services was accompanied by substantial changes in the labor market, mostly driven by demographics. Indeed, not only has a substantial population growth led to increasing the size of the labor force, but there was also a simultaneous increase in the average qualification level (Boillot 2016). The labor force participation rate in 2011-12 was 55.9% which amounted to approximately 472.9 million working individuals. Strong employment growth took place between 2009-10 and 2011-12, with a 57.2% growth in urban employment while urban residents only represented 31% of the population (ILO 2016a). Furthermore, the structural shift of the Indian economy is more impressive regarding the contribution to GDP than employment generation. The agricultural sector remains the primary source of employment for the rural workforce (62.7% of rural employment in 2011-12). Moreover, urban areas have benefitted from an increase in industry and service jobs since the 1990s, the quality of these jobs is far from meeting the requirements of employment decency[4] (Lerche 2012).

[3] Share of the population below the threshold of 1.90USD per day (PPA 2011-12).

[4] Achieving decent work, according to the ILO Declaration on Fundamental Principles and Rights at Work (1998), is based on four objectives: rights at work that follow international labor standards, employment and income opportunities, social protection and social security, social dialogue and tripartism.

2. Group characteristics and economic disadvantage in India

Contemporary history has proven time and again that economic inequality is a factor of social unrest. One type of inequality that often motivates political action and influences economic policy is *horizontal inequality* which emerges between two groups of individuals based on one specific characteristic (Stewart 2016). It has been debated whether inequality based on any group-defining characteristic reflects injustice. Local norms that are specific to a country or to a social group may shape this debate (Renaut 2014), and explain why the analysis of stigma against Black individuals is relevant in the United States labor market, why the Mapuche community suffers from exclusion in the Chilean labor market or why homosexuality is not perceived in the same way from one society to another. Indeed, "*the salience of a particular identity is likely to be accentuated by large inequalities and discrimination across […] groups*" (Stewart, 2016). Furthermore, the analysis of horizontal inequality does not necessarily require the potentially stigmatized group to be a minority in terms of population shares. Indeed, number not always being a synonym of power in a given society, a group may face unequal treatment despite representing an important share of the population.

The first group-defining characteristic that is part of our research question is *gender*. International institutions encourage assessing the extent of gender inequalities in order to design relevant public policies. Achieving gender equality and empowering all women and girls before 2030 is the fifth Sustainable Development Goal. According to The World Bank Development Report (2012), entitled "*Gender equality and development*," gender equality is an essential goal in itself, which is why development should be accompanied by narrower gender gaps in well-being. Gender equality is also an *instrument* for development. Narrower gaps in educational outcomes and access can increase productivity. Moreover, empowered women can make better decisions for their children and more representativity and inclusiveness in public institutions are likely to result in a better development path. Globally, the gender gap in primary and secondary education is narrowing. However, critical gaps remain in higher education (especially in STEM education[5]), employment rates, seats in parliament and elderly pensions. The Gender Development Index[6] of India was 0.841 in 2017 which makes the country rank

[5] STEM education includes science, mathematics, engineering, manufacturing and construction.

[6] The Gender Development Index developed by the UNDP is the ratio of the female to male Human Development Index values. As a reference, the average GDI of Low Human Development, Medium Human Development, High Human Development and Very High Human Development Countries are 0.862, 0.878, 0.957, 0.983 respectively.

131st in the world. This high level of gender inequality is led by all three dimensions of the Human Development Index, namely life expectancy at birth (67.3 years for women and 70.4 for men), mean years of schooling (4.8 for women and 5.2 for men) and gross national income per capita[7] (INR 2722 for women and INR 9729 for men) (UNDP 2017). The clear disadvantage of women in India has a labor market dimension. With one of the lowest female labor force participation rates in the world, the country saw a decline of more than 11 percentage points in female labor between 2004-05 and 2011-12. The quality of female labor is also unsatisfactory, especially in urban areas (ILO 2016a).

The second group-defining characteristic that is part of our research question includes two types of socio-religious groups based on *caste and religion*. In societies in which multiple socio-religious groups coexist, there is a risk of discrimination. The Indian society is traditionally organized in hierarchical groups identified as castes[8]. This system finds its origins in Hindu religious texts which are called *Vedas*, introduced in the Indian society during the invasion of the Aryans in 1000 B.C. The caste system divides the population into hereditary groups, which are supposed to be spatially separated from each other for instance by living in different zones (Deliège 2004) and relies on five main attributes: endogamy, hereditary membership, occupational specialization, hierarchy and commensality[9] (Klass cited by Bros (2010). *Jatis* are the most relevant caste groups and they are quite informative as to which occupation they are traditionally supposed to occupy[10]. There are more than 4500 *jatis*, which can be grouped into five hierarchically organized groups. The first four groups of *jatis* constitute *varnas* and the fifth, although also composed of *jatis*, does not constitute a *varna* (Benbabaali 2013). The groups are the following: *Brahmans* who are traditionally priests, *Kshatriyas* who occupy military and political positions, *Vaishyas* who are merchants and *Shudras* who are menial workers supposed to serve the previous three varnas. The fifth group, which does not constitute a *varna*, is called *Dalits* or "*untouchables*." This group ranks last in the hierarchy of the Hindu religion, which is built around the concept of purity. However, information on *jatis* remains quite scarce in national level datasets. Furthermore, the Indian society is not solely composed of Hindus. Many other ethnic minorities, considered as the indigenous population of India

[7] This measure is derived from the ratio of female to male wages.

[8] For a detailed presentation of the Caste system in India see Bros (2010) and (Deliège 2004).

[9] Commensality is the practice of eating together.

[10] The *jati* and the occupation it designates can be visible in surnames.

constitute a group called *Adivasis*, which is often associated with the *Dalit* group in terms of economic disadvantage.

The Indian Constitution stipulates that discrimination against specific socio-religious groups is prohibited. These groups are the Scheduled Castes (i.e. the Dalits) and the Scheduled Tribes (i.e. the Adivasis)[11]. Moreover, the same legal text lays the groundwork for one of the first affirmative action policies called "*Reservations*" which ensures a quota of SCST members in elected Seats, universities and the civil service. In 1990, reservations in publicly funded institutions were extended to a part of the Shudras called "*Other Backward Castes*" (OBC).

Islam is the second most represented religion in India. 14.2% of the population is Muslim (Census 2011). Although this community faces low economic status, it is only in 2006 that the government of Manmohan Singh decided to undertake an inquiry on the Social, Economic and Educational Status of the Muslim Community of India. The results of this inquiry are available under the name of Sachar Committee Report (Government of India 2006). The main findings of this report point out critical deprivations of the community. 38.4% of Muslims live below the national poverty line (this share is higher than that of SCSTs, who at the time had a share of urban poverty of 36.4%). According to the SCR, these differences are mostly due to low education levels (literacy or educational attainment). Intra-community disparities also exist among Muslims. In contrast to the Hindu community, where caste was originally a religious concept, Islam recognizes equality among all men. However, there is a social practice of caste in Muslim communities that is similar to the one in the Hindu community, by endogamous marriage and hierarchical organization. *Ashrafs* are the higher castes. *Ajlafs* are *Shudras* converted to Islam and *Ailafs* are untouchables converted to Islam. Both these groups are the lower castes of the Muslim community in contemporary India (Jaffrelot 2009; Government of India 2006). The Indian government recognizes Muslim Other Backward Caste groups. Indeed, among the 3,743 OBCs, 82 are Muslim. However, there is no policy to recognize them as Scheduled Castes, which would allow them to benefit more from affirmative action policies.

In this thesis, religion and caste are grouped into one variable with the following categories: Hindu Upper Caste, Hindu Other Backward Caste (or Hindu OBC), SCST, Muslim Upper Caste, Muslim OBC, Other groups (Christian, Sikh, Jain). These groups are referred to as socio-religious groups.

[11] Both groups are addressed as "SCST' in the rest of the dissertation.

3. How to analyze horizontal inequalities in the labor market?

Analyzing how gender, religion and caste can influence labor market outcomes requires concepts and tools from different branches of the socioeconomic literature. Furthermore, this literature often draws on other disciplines such as anthropology or psychology to shed light on the multidimensional nature of horizontal inequalities in the labor market.

The human capital theory, which extends the neoclassical framework[12] by relaxing some of its hypotheses, provides the most widely used tools and concepts to understand salary differentials in the labor market. Developed by Schultz (1961), Becker (1962) and Mincer (1974), this theory introduces the idea that labor supply can be heterogeneous. Facing the inability to explain wage differentials on the labor market, human capital theorists, emphasize a causal relation between individual investments in human capital and wages. Human capital is defined as productive abilities that are acquired through general and specific knowledge. Although human capital is a broad concept that can include various forms of productivity-related characteristics, its empirical analysis is restricted to observable characteristics. Mincer's model introduces the earnings function and the concept of returns to education. This function consists in estimating the effect of investment in education before entering the labor market, education acquired throughout the career (e.g. on-the-job training) and professional experience on wages. The empirical adaptability of the Mincer function makes the author's contribution a cornerstone of empirical labor market literature.

The analysis of labor supply characteristics is insufficient to understand why two equally productive individuals are remunerated differently. Extensions of the human capital theory, which analyze discrimination, focus on the demand-side behaviors that reflect the employer's taste (Becker 1971) or his perception of employee productivity in a context of imperfect information (Phelps 1972). Lang and Lehmann (2012) make a distinction between prejudice and discrimination. The first concept refers to an attitude or taste whereas the other "*refers to the treatment of people and entails treating equals unequally*." Discrimination may or may not be based on prejudice and a prejudiced employer may or may not act on its prejudice to discriminate.

[12] In a perfectly competitive market, the intersection of labor supply and labor demand, both homogenous, determine the market-clearing wage (Borjas 2010). Apart from a minimum wage that can be set by the State, this framework does not allow to understand wage differentials between individuals.

Other frameworks provide alternative explanations concerning the role of social groups in the labor market, namely the concept of *informal institution*, the *identity theory* and the analysis of *joint discrimination and intersectionality*.

Labor markets are shaped by institutions, that either act on prices or quantities, and that can be defined as a set of formal and informal rules organizing social, political and economic relations (North 1990). They constitute systems of laws, norms or conventions that influence the behaviors of both labor supply and demand by altering individual choices related to labor and pay (Boeri and van Ours 2013). The enforcement of norms related to gender, religion and caste identity qualifies as informal institutions (The World Bank 2012). Caste and gender are often considered as "*traditional institutions*" or "*religious institutions*" (Munshi and Rosenzweig (2006). In contexts such as India, the analysis of the role of these informal institutions on the labor market is all the more relevant given the low enforcement of the legal system which gives a more important role to the "*social regulation of the economy*" (Harriss-White 2010).

Sociodemographic groups reflect social constructs as well as innate characteristics. The "*natural*" or "*innate*" differences between gender groups are of physical nature and their relevance in labor market analyses rely on the idea of differential productivity related to strength. Moreover, the social psychological literature studies behavioral differentials between men and women. For instance, Harris, Jenkins, and Glaser (2006) show that women are more likely to perceive negative outcomes and are more risk-averse in domains such as gambling and health. However, they do not differ from men in social risks[13]. Labor market behavior can also be influenced by differential psychological traits of women. Babcock and Laschever (2003) show that women are less prone to engage in wage negotiations or ask for a promotion. These psychological differentials do not necessarily reflect genetic differences and their relevance may vary depending on the geographical zone of interest. "*Natural*" or "*innate*" differences can also exist between socio-religious groups if they translate an ethnic difference or if the disadvantaged group has suffered from deprivations in the long-term, leading to innate health differentials.

Furthermore, the notion of identity is key to understanding how social norms related to personal characteristics influence economic decisions. By introducing the concept of "*identity utility*," Akerlof and Kranton (2000) reintegrate the social sphere into an economic rationality framework. In their model, an individual's identity or "*sense of self*" is integrated into a general

[13] An example of social risk used by the authors is discussing opposing viewpoint with a colleague.

utility function. Economic behavior is influenced by identity in four ways. Indeed, they show that identity affects economic behavior by changing the payoffs from a person's actions and the actions of others. They also show that the inability (or difficulty) of changing identities for a person creates social exclusion from remunerative activities. However, the way a given identity is valued in a society is not necessarily rigid and identity-based preferences can change.

The simplified nature of the identity theory has however been criticized. Indeed, identities may also overlap and lead to different outcomes. Introduced by Crenshaw in 1989 to analyze the status of black women in the United States, the intersectionality approach is based on the analysis of stigma through the interaction of the characteristics that define one's identity in a given society (Halim, Yount, and Cunningham 2016; Howard Frederick 2010). Therefore, the analysis of intersectionality goes beyond the binary paradigm of discrimination analysis (someone is either discriminated or is not) and allows scenarios where someone can simultaneously be penalized and privileged (Hankivsky 2012). This approach seems to be relevant in the context of India where caste and gender both seem to have important roles to play in determining labor market outcomes. In theory, cumulating identities can range from marital status to physical appearance. A particularity in India is that caste encapsulates a large number of identities such as religious identity, social class and even skin color.

4. Combining quantitative and qualitative analyses

In order to provide empirical findings concerning horizontal inequalities across social groups, this study proposes to analyze both quantitative and qualitative data. The thesis is organized in four chapters, each focusing on a specific dimension of the labor market. Each chapter contains one research question to which we provide an answer to using a quantitative methodology as well as occasional insights from a qualitative case study.[14]

The quantitative analyses of this thesis use statistical and econometric tools to measure the extent and uncover the vectors of horizontals inequalities in India. We use data from the India Human Development Survey I and II (Desai, Vanneman, and National Council of Applied Economic Research 2012) which is a nationally representative[15] household panel database

[14] The analysis from the qualitative case study will take the form of boxes throughout the thesis. Appendix 0.1 presents additional elements on the case study.

[15] All Indian states and union territories are covered by the survey except for Andaman and Nicobar, and Lakshadweep, two union territories comprising less than 0.05% of the country's population.

collected in 2006 and 2011-2012 by the University of Maryland and the National Council of Applied Economic Research of New Delhi. Because of the substantial differences between rural and urban India, especially concerning the social stratification of the labor market, our analysis is (quasi-exclusively)[16] restricted to Urban India.

Since a national level study of urban India may insufficiently describe the local mechanisms that come into play in the labor market, we complete the findings from the quantitative analyses with insights from a field study. In the course of our Ph.D., we were given the opportunity to collaborate with the French Institute of Pondicherry to conduct a socioeconomic analysis of the labor market in Ranipet (Tamil Nadu)[17]. Observing the local specificities of the labor market through the eyes of labor market actors is an interesting addition to the trends and mechanisms observed at the national level. This methodological exercise of confronting quantitative empirical findings to the perceptions of individuals contributes to our research for the following reasons. First, a local approach is informative on the extent to which individuals place a high value to our problematics of interest. Moreover, factors we have not considered as important may arise as relevant factors in the shaping of labor market outcomes. Finally, a local approach may help us validate or nuance specific quantitative findings. Given the nature of qualitative data, we analyze what we have seen on the field using a constructivist approach, which consists in interpreting the meaning of individual and collective actions and behaviors. Such reasoning requires departing from a neutral standpoint and engaging our opinion and own perceptions as a researcher.

The fieldwork took place in Ranipet, a town of 51,000 inhabitants, located in the Vellore district of the State Tamil Nadu. This State is one of the most urbanized ones in India with 48.4% of the population located in urban areas. It is also relatively developed. With public policies that brought the State to the second rank regarding universal schooling, the commitment towards inclusiveness has nevertheless been insufficient to erase gender and caste discrimination in Tamil Nadu (Vijayabaskar 2004). Ranipet, which presents characteristics of the diffused form of urbanization that is taking place in the State (Denis and Marius-Gnanou 2010), is part of a Special Economic Zone and is a manufacturing center that produces leather goods for the

[16] The only exception is the analysis presented in the Section 2 of Chapter 1. The inclusion of the rural sample was necessary for statistical purposes, in order to allow for sufficiently large samples.

[17] This study is part of an ongoing research project with K. Marius and G. Venkatasubramanian.

domestic and foreign markets. The local economy is built around the leather industry as it provides manufacturing and other forms of employment.

The fieldwork took place between 2015 and 2016. Semi-directive interviews of workers were conducted with the purpose of maximizing the diversity of profiles such as gender, caste or occupation type.[18] This rich information source on the labor market of Ranipet constituted the base of our observations and allowed us to gather information not only on the labor market configuration but also on the perception of the place discrimination and inequality hold in the labor market.

5. Research question and outline of the thesis

The main research question of this thesis is as follows: **what are the mechanisms and the extent of horizontal inequalities based on gender, religion and caste in the urban Indian labor market?** We address the main question in a four-pronged approach which analyzes (i) the characteristics of individuals who do not engage in the labor market as well as the stakes of female labor market exclusion on children's education, (ii) the different patterns of labor market mobility, (iii) the forms of labor market segmentation in a predominantly informal labor market and, (iv) wage gaps and potential labor market discrimination. The study provides an analytical contribution to the literature on the urban Indian labor market. The concept of horizontal inequalities is incorporated in the study through the analysis of heterogeneity in the mechanisms observed. This study contributes to the socio-economic literature either by providing insights on issues that remain scarcely analyzed in India (e.g. seclusion of women in the labor market) or by updating well-established problematics with more precise specification or alternative methods. The systematic separation of the Muslim group between a Muslim Upper Caste and Muslim OBC group, for instance, provides renewed insights on labor market mechanisms that have been analyzed by completely ignoring the Muslim group or by considering it as a homogeneous.

The thesis is organized as follows.

Chapter 1 analyzes the determinants and the intergenerational consequences of labor market exclusion. In the first section, we define labor market exclusion as an alternative concept that encompasses unemployment and inactivity. Then, we provide an estimation of the determinants

[18] Additional information of the field work is provided in Appendix 0.1.

of labor market exclusion with a focus on the role of religion, caste and gender. The results indicate direct effects of caste, religion and gender on the probability of being excluded from the labor market. Several indirect links are also found, specifically when combining these sociodemographic groups to age or education level. Given the important exclusion of women from the labor market, the question of the persistence of this phenomenon in time is then explored in the second section. By detecting the gender-specific consequences of a mother's labor market status on children's education, we can provide insights on how gender attitudes are transmitted throughout generations.

Chapter 2 proposes to detect transitions in the medium-run between 2005 and 2011-12 and analyze their determinants. We observe the magnitude of movements for different gender, religion and caste groups. We contribute to this literature by offering a labor market perspective which combines an analysis of career mobility (namely between casual and regular occupations, industries and skill levels in occupations) and hourly earnings mobility. In comparison to measuring intertemporal household income mobility, the analysis of intertemporal rank change in the hourly earnings distribution provides information on how labor markets are a vector of social mobility. Moreover, focusing on workers' hourly earnings has a practical appeal as it allows to compare patterns of mobility between men and women. We provide a detailed description of patterns of mobility. We also identify its determinants by considering several potential econometrical issues such as sample selection, measurement error and the endogeneity of initial earnings.

Chapter 3 explores the heterogeneity of the urban Indian labor market. We consider two different sectors of the labor market: household businesses (which include own-account workers that may benefit or not from contributing household members and more formal small enterprises) and salaried work. We analyze the potential heterogeneity of both sectors by examining how many types of earnings structure they contain using a semi-parametric approach. The results allow us to address the following questions: (i) Can we identify a duality that could relate to formal versus informal duality in one of the sectors? (ii) is there an opportunity versus necessity form of duality in this predominantly informal labor market? (iii) How do social identity variables, namely gender, caste and religion, come into play in the segmentation process?

Chapter 4 proposes to explore wage gaps on the grounds of gender, religion and caste by using multiple decomposition methods. The chapter contains two main types of analyses: wage decompositions at the mean and across the distribution. First, we explore wage gaps at the mean using parametric decomposition methods that account for the sample selection bias. We compare findings from the parametric decomposition to those from a non-parametric decomposition method developed by Ñopo (2008) which allows comparing wages between matched individuals. We then proceed to the descriptive analysis of wage gaps across the distribution by comparing the matched samples generated by the Ñopo decomposition method.

Chapter 1. Premarket factors and labor market exclusion in the Indian labor market

Introduction

The lack of choice regarding work can take many forms ranging from forced labor to being forbidden from working. These situations constitute significant breaches to freedom and seem to be a cause for concern in India. On the one hand, bonded labor practices are still frequent (mostly in rural areas)[19] and on the other hand, the large share of inactive individuals indicate possible mechanisms of more or less voluntary exclusion from work, which this chapter proposes to explore.

The demographic dividend due to the booming Indian population could present an economic opportunity in terms of growth possibilities and poverty alleviation. However, for India to reap the benefits of its demographics, the labor market needs to be able to absorb all potential workers. Employment statistics show that much remains to be done, especially for women. Along with a 3.4% unemployment rate in 2017 according to the World Employment Social Outlook (ILO 2017c) and substantial inactivity rates[20] (ILO 2016a), the large share of part-time employment shows missed productive opportunities at the macroeconomic level and considerable vulnerability at the microeconomic level.

The working-age population that remains outside of the labor market can either be unemployed or inactive. In this chapter, we use the term "*labor market exclusion*" as a concept that encompasses the voluntary (i.e. self-exclusion) and involuntary (i.e. unemployment due to a lack of labor demand) nature of not being actively occupied. In other words, *labor market exclusion comprises the inactive and unemployed individuals*.[21] Considering the sociodemographic dimension of labor market exclusion is particularly relevant in India where

[19] For an analysis of bonded labor practices in India, see for instance Guérin (2013) and Guérin, Venkatasubramanian, and Kumar (2015).

[20] Labor market participation rates do not exceed 63.7% between 2004 and 2012 (ILO 2016a).

[21] In the literature, the union of these two groups does not have a specific term, it is usually addressed as "unemployed and inactive" or "people outside of the labor market".

substantial inequalities are observed across gender, religion and caste groups. Both types of exclusion have a self-fulfilling dimension because of persistence in time. First, religion-based or caste-based labor market exclusion is bound to maintain specific types of societal inequality as it limits vulnerable groups' opportunities for climbing the social ladder. The caste system in India specifically enters this dynamic because of its endogamous nature. Moreover, the exclusion of women from the labor market ensures her economic dependency to the person who earns money in the household, which in most cases is her father or husband, but also vehiculates a specific mindset concerning gender attitudes throughout generations.

This chapter aims to analyze the determinants and the intergenerational consequences of labor market exclusion. In the first section, we define labor market exclusion as an alternative concept that encompasses unemployment and inactivity. Then, we provide an estimation of the determinants of labor market exclusion with a focus on the role of religion, caste and gender. We analyze direct as well as indirect associations (i.e. that are mediated by disparities in premarket factors) between group membership and labor market outcomes. The results indicate direct associations between the probability of being excluded from the labor market and the sociodemographic groups. Several indirect associations are also found, specifically when combining the groups to age or education level. Given the important exclusion of women from the labor market, the potential persistence of this phenomenon in time is then explored in the second section. By detecting the gender-specific consequences of a mother's labor market status on children's education, we can provide insights on how actual labor market status contributes to future educational gaps and to the transmission of gender attitudes throughout generations.

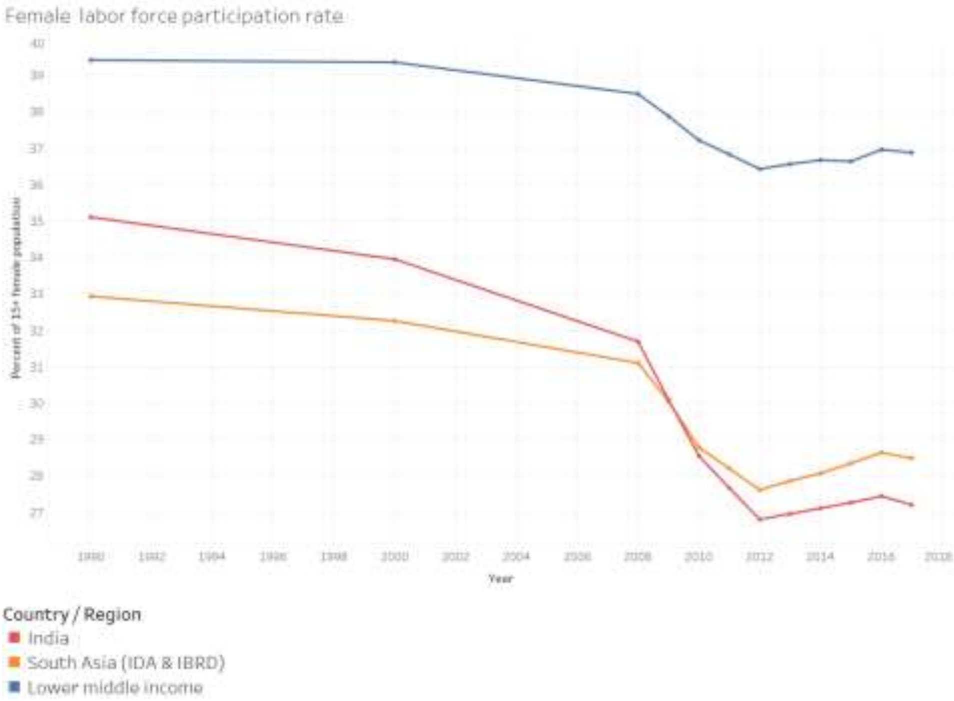
Section 1: Premarket factors and labor market exclusion

The current demographic structure of India could present a sustainable solution for growth and development (Mitra and Verick 2013). However, in order for an increasing share of the working age population relative to the total population to be beneficial in terms of economic development, some conditions relative to labor supply and demand must be met. The link between age structure and growth is mainly driven by improvements in the population's educational attainment (Renteria et al. 2011; Crespo Cuaresma, Lutz, and Sanderson 2014). Moreover, the inability for a country to provide employment in terms of quantity and quality may also curb growth potentials. Although the 34 Indian States have benefited from the demographic dividend, there is no correlation between the States with higher shares of working age population and per capita income growth (OECD 2017). Structural factors such as educational attainment (OECD 2017) and health (Bhattacharya and Haldar 2015) may be the reason why India is not able to economically benefit from its demographics. In fact, the country faces significant shares of non-utilized or under-utilized labor supply.

Whether unemployment and inactivity are driven by demand or supply characteristics, identifying which individuals are more likely to be in these categories is essential for improving public policy design and targeting. In particular, it is important to assess the extent of horizontal inequalities among different socio-demographic groups concerning employment opportunities. In the case of India, significant gaps in human development and human capital seem to persist across gender (Klasen and Pieters 2015), religion (Bhaumik and Chakrabarty 2009) and caste groups (Thorat and Newman 2010).

A majority of Indian women does not participate in the labor market and the Female Labor Market Participation (FLMP) rate is much lower in India than in other Lower-Middle-Income Countries (Figure. 1.1). Although the declining trend of FLMP since the 1990s is not an Indian specificity, a steeper decline compared to the South Asian average is observable since 2006 because of a lack of employment creation that matches women's qualifications (Chaudhary and Verick 2014). The relationship between economic development and FLMP is complex. Since the seminal study of Goldin (1995) in which she demonstrates that FLMP and development follow a U-shaped curve, many studies have further analyzed the mechanisms leading to this trend. Throughout a country's development, the reasons for an initially high participation rate of women is most probably driven by necessity, because of low levels of household income and lacking systems of income protection such as unemployment insurance (Cazes and Verick 2013). As the level of development increases and leads to larger per capita incomes, women withdraw from the labor market. Finally, when a country is developed, FLMP increases because of higher educational attainment, changing norms, increased labor demand, etc. (Chaudhary and Verick 2014). Moreover, there is an emerging consensus on the fact that increased FLMP can contribute to economic growth (Esteve-Volart 2004), to decreasing gender inequality and increasing women's empowerment (Kabeer 2012; Mammen and Paxson 2000).

Figure 1.1. Female labor force participation rate (1990-2017)



Source: Author’s illustration from the World Bank Indicators Database

Social norms dictate family-related choices and employment decisions both at the supply side (e.g. choosing not to work for child-rearing purposes) and at the demand side (e.g. employment discrimination). These norms provide an explanation as to why women remain outside of the labor market. India being at an ambiguous development stage, with important growth levels but unequal redistribution (both between and within households), the mechanisms of FLMP are likely to be affected.

Socio-religious groups are potential determinants of labor market exclusion as well. The cultural link between caste and career paths is inherent to this occupation-based endogamous system which has been enforced for many generations regardless of the modernization of the economy (Munshi and Rosenzweig 2006). However, it might also affect labor market exclusion for reasons such as employment discrimination towards lower castes (Deliège 2004). Furthermore, religion is also a traditional institution which can be associated with discriminative behaviors, directed for instance towards the Muslim community (Karimullah and Kalpagam 2010).

The ways in which these socio-demographic factors influence labor market participation are not necessarily direct. Premarket factors regarding education and health have a potentially important role in keeping individuals out of work. These premarket inequalities may also reflect how individuals indirectly suffer from forms of discrimination outside of the labor market, in terms of education or health for instance.

In this section, we provide a detailed analysis of labor market exclusion by exploring the direct and indirect correlations between labor market exclusion and religion, caste and gender. We estimate the probability of labor market exclusion using a multinomial logistic regression. This analysis allows us to establish profiles of individuals based on their gender, religion and caste. This study considers various dimensions of premarket inequalities between groups. It also provides a joint analysis of gender and religion/caste and highlights the diversity of situations. By combining these group variables to other factors such as age or education, we provide detailed evidence on the dynamics of labor market exclusion.

The rest of this section is organized as follows: After presenting the concept of labor market exclusion and the literature used to frame the analysis (1), we provide a descriptive analysis of premarket horizontal inequalities and labor market exclusion (2) and present the methodology and results of the empirical analysis (3) before proposing a discussion (4).

1. Premarket discrimination and labor market exclusion in developing countries: concepts and literature

After defining labor market exclusion (1.1), this section discusses the possible theoretical channels of labor market exclusion (1.2) and reviews the literature on the forms of horizontal inequalities between group characterized by gender, religion and caste (1.3).

1.1. Concepts of labor market status and labor market exclusion

Labor market participation does not only entail individual decision. Institutional factors, formal ones as the legal framework or informal ones related to the social regulation of the labor market, are also potentially related to labor market outcomes. A challenging issue is the identification and characterization of individuals who are not occupied in the labor market, particularly when the rate of unemployment is marginal compared to the rate of inactivity.[22] The delimitation of unemployment and inactivity as it is defined by the ILO is based on the "*willingness to work*" and "*active research*." However, in many contexts, willingness is either hard to identify by the policymaker or even by the individual concerned. It is plausible to find situations where an

[22] As pointed out previously, the unemployment rate in India was about 3.4% in 2017 (ILO 2017c) and the inactivity rate was about 36% between 2004 and 2012 (ILO 2016a).

individual does not want to work but, but facing pressure by family members or peers to do so, claims to look for a job. Hence, this individual can be considered as unemployed rather than inactive in survey data. The reverse situation is also plausible. For instance, in cases of discouragement, an individual can claim giving up job search despite remaining registered as a job-seeker and receiving information on potential employment opportunities.

The economic literature usually addresses the union of the unemployed and the inactive by juxtaposing both terms or by qualifying individuals as "*those who do not work*." Nevertheless, we argue that the concept of exclusion is adequate to qualify the group as a whole, while allowing to apprehend the diversity of situations for individuals who do not work. Thorat and Newman (2010) use this concept when addressing economic discrimination faced by lower castes. They borrow the term "*exclusion*" from Sen's analysis of social exclusion, which is a broad and relative concept used to address deprivations in various aspects of individual and social life (Sen 2000).[23] Relative to labor market participation, he proposes two different concepts: unfavorable exclusion (i.e. excluding someone from a given situation) and unfavorable inclusion (i.e. including others in a given situation). Moreover, the limited extent of one's capabilities can lead to self-exclusion. Based on these concepts, the term "*labor market exclusion*" seems to comprehensively qualify those who do not take part in the labor market in India. Indeed, an individual who is outside the labor market is either excluded (by the employer or by the lack of labor demand) or self-excluded. The latter category includes individuals who choose not to work because of a preference for leisure and those who restrict themselves from working for other reasons. Moreover, this categorization is practical for empirical analyses since many data sources do not contain information about the voluntary nature of being out of work.

The Indian labor market being largely composed of workers with part-time or seasonal jobs, it is important that the concept of labor market exclusion account for the *partial exclusion* these activities represent. To do so, we distinguish *complete labor market exclusion* from *partial labor market exclusion*. *Complete labor market exclusion* concerns individuals who willingly or unwillingly do not take part in the labor market for a given period. If we consider the usual terms to identify the working-age population, this group comprises the unemployed and the

[23] Social exclusion as defined by Sen (2000) concerns individuals who are denied "a livelihood, secure, permanent employment, earnings, property, credit or land, housing, consumption levels, education, cultural capital, the welfare state, citizenship and legal equality, democratic participation, public goods, nation or dominant race, family and sociability, humanity, respect fulfilment, and understanding". For a detailed analysis of this concept and its links to economic discrimination see Thorat and Newman (2010).

inactive. According to the definition of the ILO in 1954, a person is considered unemployed if this person is not working, available for work, and looking for work. A person who is not working but either not available or not looking for work is called an inactive person. *Partial labor market exclusion* concerns all types of time-related under-employment. The 16th International Conference of Labour Statisticians of 1998 adopted a resolution that established the most commonly used definitions of underemployment and inadequate employment situations. Time-related underemployment identifies individuals who are active in the labor market, are willing to work more hours and are available to do so. These individuals are only considered to be underemployed if the number of hours they work is lower than a given threshold. Partial labor market exclusion also includes voluntary part-time labor for the same reasons as the inclusion of voluntary inactivity in the category of complete labor market exclusion.

Table 1.1. Concepts of labor market exclusion

Category	ILO equivalence	Nature
Complete labor market exclusion	Unemployment	Involuntary (lack of demand or discrimination)
	Inactivity	Voluntary (leisure or self-exclusion)
Partial labor market exclusion	Time-related underemployment	Involuntary (lack of demand or discrimination)
	Voluntary part-time employment	Voluntary (leisure or self-exclusion)

Source: Author

The other form of underemployment pointed out by the ILO (i.e. inadequate employment situations) is commonplace in developing economies. These situations involve individuals who are active in the labor market but are unsatisfied with their current employment situation being unable to use their capacities at their job fully. In this case, the workers must be willing to change their employment situation because of inadequate or insufficient use of their skills, inadequate income, or excessive hours of work. Although relevant in the analysis of labor in India, this category falls out of the ambit of this study.

1.2. Channels of labor market exclusion: premarket factors, discrimination and self-exclusion

Theories about unemployment, whether they stem from the neoclassical theory or the Keynesian theory, only partially cover the reality of developing labor markets. Indeed, many other configurations may be more relevant such as underemployment, seasonal employment, forced or quasi-forced employment (i.e. debt bondage), forced unemployment, etc. Moreover, explaining inactivity solely by the preference of leisure over work also seems insufficient. High levels of legal non-compliance to minimum wages and the existence of an informal economy make the theories linking unemployment to a minimum legal wage or unemployment benefits unfit to describe these labor markets. In this specific setup, informal institutions such as gender or ethnicity seem to have important roles to play in the determination of labor market exclusion through three channels: premarket factors, discrimination and self-discrimination.

"*Premarket factors*" (also addressed as "*non-market factors*" in the literature) is a concept used in labor economics to identify a vast array of characteristics that individuals hold before they enter the labor market. The concept refers to individual characteristics that influence labor market outcomes and that have been acquired outside the labor market. Carneiro et al. (2005) consider that a premarket factor is "*not affected by expectations or actual experiences of discrimination in the labor market*." Although premarket factors can include many variables, studies often use education input (years of education) or education output (test scores) to account for skill differentials between groups (Neal and Johnson 1996). Inequality in premarket factors between two groups can either be caused by discrimination (e.g. differential treatment of children by teachers) or by unequal levels of physical, human or social capital. Expectations and aspirations can also alter household choices in terms of human capital accumulation. Although it is unlikely that younger children have themselves formed expectations about the labor market, parental choices can be influenced by their perception of the labor market. For instance, in the case of India, Jensen (2012) finds that when presented with more labor market opportunities, young women aspire to have fewer children. This has potential implications in terms of per capita spending in the household and can lead to premarket inequality among children from different socio-religious groups.

Labor market exclusion can also be caused by employment discrimination. First, discrimination can lead to labor market exclusion if an individual is not hired and stays unemployed because of employer taste. In this case, the employer chooses not to employ an individual, who is equally productive to others, by showing an irrational behavior linked to his preferences. Moreover, in the presence of imperfect information, statistical discrimination can lead the employer to rationally choose not to employ an individual, who is equally productive to others.[24] This transmission channel most closely reflects the idea of involuntary exclusion. However, job discrimination is more likely to lower the reservation wage of individuals facing prejudice (Lang and Lehmann 2012) than to lead them to abandon job search. Furthermore, in a context where individuals often fall back on informal self-employment, this type of discrimination may not significantly impact the share of individuals who stay completely excluded from the labor market. Nevertheless, a channel that potentially contributes to the inflation of this share is the existence of the *self-exclusion* or *self-discrimination* where one's perception of the labor market influences decisions to participate or not in the labor market. From that point of view, women can decide not to participate in the labor market if they consider that their role in society is limited to carrying children and being primary caregivers. In terms of religion or caste, self-exclusion is more likely to cause individuals to occupy specific occupations rather than excluding them from the labor market completely.

1.3. Horizontal/group inequalities and labor market exclusion in developing economies

Belonging to a specific group can make an individual more prone to being excluded from the labor market because of the mechanisms described above. This section explores three types of group inequalities relevant to the Indian labor market: gender, religion and caste.

1.3.2. Female labor participation in developing countries

Household dynamics play a big role in the determination of FLMP as women have to choose between market work, home production and leisure whereas men only choose between market work and leisure (Mincer 1962). According to Goldin (1995), economic development and FLMP follow a U-shaped relationship. The decision of women's labor supply is driven by two main factors: the opportunity cost of her time and household income, often represented by the spouse's wage. Initially, when a country has very low development levels, the importance of agricultural activities is correlated with important FLMP, which can be paid or unpaid labor. Then, as the country develops, a decline in FLMP is led by a shift from agricultural, household or small-scale production to other industries, thus creating an income effect (i.e. a woman has

[24] Chapter 4 presents the concepts of labor market discrimination.

to work fewer hours since the husband earns more) reinforced by a small substitution effect (i.e. a change in hours of work in respect to a change in a woman's own wage). In later stages of the development process, as female education attainment increases leading to better access in higher-paying occupations, the income effect becomes less important and the substitution effect increases, leading to an increase in FLMP. This pattern is reinforced by other social and cultural factors such as a decrease of stigma towards working women, a decrease in fertility rates, a change in household dynamics in which men and women share household and market responsibilities, or an increase of divorce rates (Blau and Kahn 2017; Goldin 1995; Mammen and Paxson 2000). The FLMP rate in India was less than 30% in 2016, which is inferior to other Lower Middle-Income countries whose average FLMP was about 37%. Since 2009, this level has also become smaller than the average of South Asian countries (ILO 2017c). Klasen and Pieters (2015) show that the low and declining FLMP between 1987 and 2011 in India is caused by multiple factors. First of all, education is associated with an increased utility cost of engaging in low-skilled work. Once women access higher education, this utility cost disappears because of increased access to white-collar jobs. Thus, education and FLMP follow a U-shape curve probably caused by social stigma towards low-skilled jobs. Moreover, Lahoti and Swaminathan (2016) reject Goldin's hypothesis of a U-shaped relationship between FLMP and economic development. They show that the economic growth of Indian States between 1983 and 2012 has not been labor-intensive and that India's growth was not led by agriculture and manufacturing. As a consequence, women who were more concentrated in these sectors did not benefit from an increase in employment as they are not qualified for jobs in more "*growth-leading*" sectors. Therefore, the emergence of the Indian economy might paradoxically influence women's decision to exit the labor market because of relatively improved economic situations.

Social norms regarding female work also influence FLMP (Chen and Drèze 1992) by affecting both labor demand and supply.[25] The stigma against women's work can lead employers to discriminate women by not hiring them. This type of discrimination is more likely to reallocate female labor into other occupations (whether salaried work or self-employment) rather than directly lead a woman to abandon job search in the short term. Moreover, employment discrimination and wage discrimination against women reflect the general opinion against

[25] We review the literature on social norms and female labor in Section 2 of this chapter.

female labor and might lead to self-exclusion, especially through the perpetuation of gender norms.

1.3.3. Horizontal inequalities between socio-religious groups

The channels through which religion and caste can affect the probability of entering the labor market can be classified into two groups: direct and indirect channels.

Among the direct channels, two main factors can be listed. Since occupational specialization is an important feature of the caste system, being a member of a specific caste group is likely to influence occupational choice (Gang, Sen, and Yun 2017). Second, there also is a possibility of employment discrimination from employers (Thorat and Attewell 2007). Nevertheless, as for gender, employment discrimination may only slightly influence the probability of labor market exclusion and lead someone to apply to other jobs rather than abandoning job search altogether. Symmetrically, behaviors such as nepotism[26] can benefit workers belonging to specific castes making them more likely to get a job even if they are equally productive as another worker, or if they are less productive. Employment discrimination against Muslims in India has received little attention in the literature (Karimullah and Kalpagam 2010). Thorat and Attewell (2007) show the existence of social exclusion of Muslims at the first-stages of recruitment in the formal sector. Conversely, Banerjee et al. (2009) find no proof of employment discrimination against Muslims in call-centers and the IT sector of Delhi. Both studies are based on experiments in which the authors send false applications to job advertisements. Despite their interesting findings, the same type of study cannot be conducted in the informal economy which constitutes a large share of employment.

The indirect channels through which religion and caste can influence the probability of working are linked to endowments. Indeed, differentials in various types of capital held by specific socio-religious groups can create unequal access to employment. SCSTs are particularly disadvantaged compared to the rest of the population concerning education (Mehrotra 2006; Halim, Yount, and Cunningham 2016). Concerning health outcomes, Borooah (2012) shows that there is a social gradient of health in India. Focusing on death rates, prenatal and postnatal care and the health situation of the elderly, he finds that groups higher up the social ladder have better health outcomes. Compared to Hindu women, Christian and Muslim women are less

[26] Nepotism can be defined as the set of behaviors that consist in preferring a group to another (Borjas 2013).

likely to receive prenatal care. Moreover, compared to Hindu Upper Castes, the average age of death is significantly lower by 4.9 years for STs, 7.1 years for SCs and 6.1 years for Muslims.

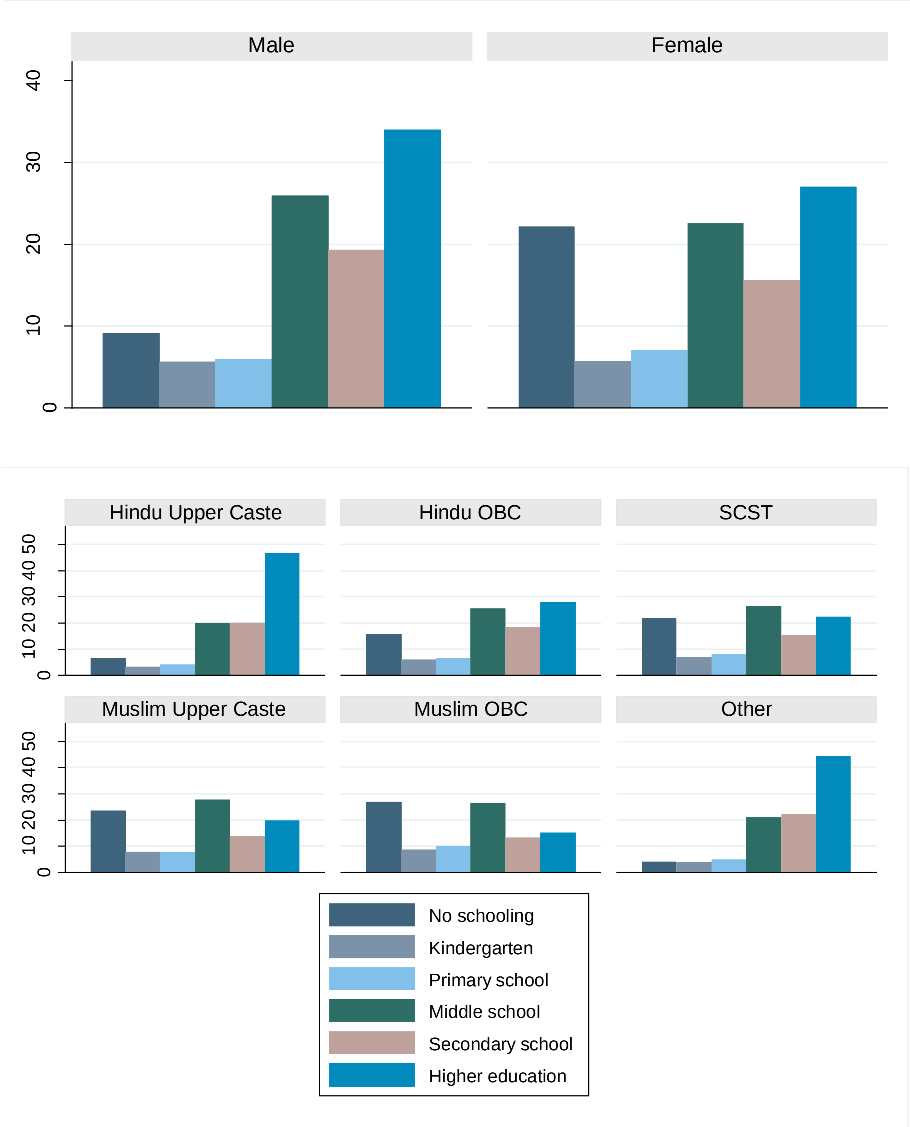
2. A descriptive analysis of premarket inequalities and labor market exclusion in India

This section presents a descriptive analysis that allows observing the extent of premarket inequalities in India as well as the differences in the characteristics of the working age population across different labor market statuses. The following empirical analysis uses the second wave of the nationally representative IHDS data (2011-12). The sample is composed of all interviewed individuals who are between 15 and 65 years old, living in urban areas and who are not currently enrolled in an educational institution.

2.1. Group inequality in premarket factors

The aim of this first step of the descriptive analysis is to point out the existence and the extent of horizontal inequalities concerning two sets of premarket factors, education and health, in the working age population of India. Literacy and educational attainment are two relevant indicators of premarket inequalities. Differentials across two groups can be caused by economic factors (e.g. different levels of physical or social capital) or by discrimination. Although, we cannot disentangle one from the other, looking at children's perception regarding schooling provide us with interesting information on potential discrimination and preferences.

Figure 1.2. Highest educational attainment by gender, religion and caste



Source: Author's calculations from IHDS (2011-12)
Note: The y-axis refers to the shares of different educational attainment levels per subsample.

Figure 1.2 shows the percentage of different educational attainment levels per gender, religion and caste group and Appendix 1.1 shows the adult literacy rates per group. Considerable gender differences are visible.[27] In the working-age population, 26.2% of women are illiterate compared to only 15.6% of men. In terms of educational attainment, we can observe an inverting trend between genders: women are more present in the lowest levels of education (from no schooling to primary school) whereas men are more present in the highest ones (from middle school to higher education). Moreover, 24.4% of women are without any education

[27] Appendix 1.2 and 1.3 show the detailed results and chi-square tests.

against 9.8% of men, whereas 24.7% of women have a higher education against 32.5% of men. These large gaps are probably due to different parental choices in the education of girls compared to boys, which can take the form of different levels of investment in education (Lancaster, Maitra, and Ray 2008; Azam and Kingdon 2013). Moreover, it is plausible that the behavior of teachers differs depending on whether they teach boys or girls. This may be more visible in coeducation schools which does not represent the majority of cases in India.

The results concerning caste also reflect important horizontal inequality, the highest share of illiteracy is observed for Muslim OBC (31.7%) and Muslim Upper Castes (27.1%). Conversely, the highest percentages of literacy are observed for Other groups (i.e. Christians, Jains and Sikhs) and Hindu Upper Castes. 43.6% of Hindu Upper Castes have a higher education level, whereas only 26% of Hindu OBCs, 21% of SCSTs, 18.4% of Muslim Upper Castes and 13.9% of Muslim OBC reach this level.

These observed differentials in premarket factors can be the consequence of two main issues: sociocultural and economic differentials between groups may lead to different levels of human capital and specific groups may face discrimination. If we consider the first possible cause, the mechanisms by which women and religion or caste groups are affected are different. Economic gender inequality that can cause differentials in human capital pertains to intra-household income allocation, which is difficult to observe in our data. Nevertheless, capital levels significantly differ between socio-religious groups. For instance, if we consider a total household asset index[28] that ranges from 0 to 30, Table 1.2 shows that Hindu Upper Castes have a significantly higher score (22.43) than the other groups (19.42) at the 1% level.

Table 1.2: Asset score by religion and caste group

Group	Asset score average	Asset score median
Hindu Upper Caste	22.43	23
Hindu OBC	19.91	20
SCST	19.68	19
Muslim Upper Caste	18.60	19
Muslim OBC	18.96	19
Other	23.47	24

Source: Author's calculations from IHDS (2011-12)
Note: Student t-tests show that the difference between the average of each group compared to others is significant at the 1% level.

[28] The index represents the total of the following owned assets for each household: sewing machine, mixer/grinder, motor vehicle, TV, air cooler, air conditioner, electric fan, chair/table, cot, telephone, cell phone, refrigerator, pressure cooker, any vehicle, car, clock/watch, washing machine, computer, credit card, two clothes, footwear, piped indoor water, separate kitchen, flush toilet, electricity, solid wall (mentioned as *Pucca* in the IHDS questionnaire), roof and floor.

Discrimination faced by individuals in school can also alter one's ability to acquire human capital. Teacher bias can be very hard to detect although in India it can take enormous proportions.

Box 1.1. An experience of discrimination towards *Dalits* in schools

During our fieldwork, a *Dalit* music teacher described her father's experience of discrimination while he was a primary school student in Southern Tamil Nadu. In her opinion, discrimination against the untouchables is a common occurrence in schools. Teachers, who are usually from the *Brahmin varna* tend to neglect *Dalit* children in very explicit ways. Some of the teachers, unwilling to share the same space as a *Dalit* child, would prevent them from entering the classroom. In this case, her father was supposed to attend the class from outside. He was given paper and pencil to do his exercises outside. Being an eager-to-learn individual, her father completed school despite facing multiple occurrences of discrimination. He then succeeded the Civil Service Examination and became a public servant. Her father's experience motivated her to become a music teacher. She also benefitted from physical and social capital because of her father's income, more importantly, she claims that she inherited from her father's motivation.

Despite facing severe forms of discrimination, this man's experience points out that more effort is required for a *Dalit* child to succeed in school. Although this illustration shows that discrimination can be overcome, this type of path is probably rare in contemporary India. Furthermore, this experience shows that the difficulty of learning may encourage a selection effect: children who have the ability to complete school despite additional obstacles are probably children who present better chances of succeeding than others in any case. They may be more motivated or have higher learning abilities than the average.

This experience also shows a form of institutionalization of exclusion in which *Dalit* children are not allowed to share the same space as the other children, but they are still members of the class, as acknowledged by the teacher who let them listen from outside of the class, do the same homework and take the same tests as others. Prejudice against SCST children in school can sometimes take more extreme forms than the one described above. For instance, SCST children are often humiliated by other children since they are considered as less intelligent (Hoff and Pandey 2014).

Source: Author

Although information on school discrimination is unavailable for the adult sample in the dataset, interesting variables concerning children's educational experience are provided. More precisely, the children have been asked whether the teacher was nice, fair, good and biased.

Table 1.3 shows that girls perceive that their teachers are biased more often than boys. The distribution between both gender groups is significantly different at the 10% level. However, they think that their teachers are nice more often than boys. It should be noted that schools in India are mostly same-sex, which implies that the described perceptions are not necessarily shaped by the way the other gender is treated by the teacher.

Table 1.3: Children's perception of education by gender

	Male	Female	Total
Is the teacher nice?			
Nicely	82.2	84.2	83.2
Somewhat nicely	16.3	14.5	15.4
Not nicely	1.5	1.3	1.4
Total	100.0	100.0	100.0
	Pearson chi2(2) = 8.303; P-value = 0.016		
Is the teacher fair?			
Rarely/Never	86.0	86.8	86.4
Sometimes	10.4	9.7	10.1
Often	3.5	3.5	3.5
Total	100.0	100.0	100.0
	Pearson chi2(2) = 2.134; P-value = 0.344		
Is the teacher good?			
Excellent	29.5	28.4	29.0
Good	66.4	68.0	67.2
Fair	3.7	3.4	3.5
Poor	0.4	0.3	0.3
Total	100.0	100.0	100.0
	Pearson chi2(3) = 5.472; P-value = 0.140		
Is the teacher biased?			
Rarely/Never	91.2	90.0	90.6
Sometimes	6.6	7.4	7.0
Often	2.2	2.6	2.4
Total	100.0	100.0	100.0
	Pearson chi2(2) = 5.664; P-value = 0.059		
Does the child enjoy school?			
No	3.0	2.4	2.8
Yes	97.0	97.6	97.2
Total	100.0	100.0	100.0
	Pearson chi2(1) = 4.113; P-value = 0.043		

Source: Author's calculations from IHDS (2011-12)

Concerning religion and caste perception differentials (Table 1.4), the results are much more polarized. All the chi-square tests are significant at the 1% level which indicates that a child's experience in school differs by religion or caste group. SCST and Muslim children perceive that their teacher is biased more often than other children. Note that these perception differentials do not necessarily translate actual inequalities. Indeed, it is possible that mindsets and the expectations of children from teachers differ from one group to another. Disparities in school-related household expenditures may induce a heterogeneity in well-being in school, leading to different perceptions.

Table 1.4: Children's perception of education by religion and caste groups

	Hindu Upper Caste	Hindu OBC	SCST	Muslim Upper Caste	Muslim OBC	Other	Total
Is the teacher nice?							
Nicely	86.3	83.4	81.4	83.5	82.2	83.1	83.2
Somewhat nicely	12.7	15.2	17.0	14.9	16.7	16.4	15.5
Not nicely	1.0	1.4	1.6	1.6	1.1	0.5	1.4
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Pearson chi2(10) = 26.964; P-value = 0.003							
Is the teacher fair?							
Rarely/Never	88.9	85.9	85.3	85.8	87.4	92.7	86.5
Sometimes	8.0	10.1	11.5	9.8	10.0	3.9	10.0
Often	3.1	4.1	3.2	4.4	2.7	3.4	3.5
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Pearson chi2(10) = 43.050; P-value = 0.000							
Is the teacher good?							
Excellent	29.7	29.9	25.4	35.3	28.8	42.1	28.9
Good	67.2	67.0	70.5	60.7	64.2	54.1	67.2
Fair	2.9	2.8	3.9	3.5	6.5	3.9	3.5
Poor	0.3	0.4	0.3	0.5	0.5	0.0	0.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Pearson chi2(15) = 109.225; P-value = 0.000							
Is the teacher biased ?							
Rarely/Never	93.6	90.3	89.8	88.8	89.9	96.1	90.7
Sometimes	4.6	7.2	8.4	6.2	7.6	2.2	7.0
Often	1.7	2.5	1.8	5.1	2.5	1.7	2.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Pearson chi2(10) = 83.930; P-value = 0.000							
Does the child enjoy school?							
No	1.6	2.2	3.7	3.1	4.0	1.0	2.8
Yes	98.4	97.8	96.3	96.9	96.0	99.0	97.2
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Pearson chi2(5) = 36.842; P-value = 0.000							

Source: Author's calculations from the IHDS data

Appendix 1.4 shows that the differences between gender, religion and caste groups concerning the numbers of days of illness are much less important than for education. However, the chi-square tests point out significant differences in both cases. The share of women to report not being ill in the month preceding the survey is 84.5% whereas the share of men who report being ill is 90.4%. More women (1.7%) report being ill for more than two weeks compared to men (1%). In terms of religion and caste, we observe that Muslim OBCs and Muslim Upper Castes have higher percentages of more than two weeks of illness. Results concerning hospitalization

(Appendix 1.5) show that long-term hospitalization is important among Hindu OBCs and very rare among Muslim Upper Caste individuals. Note that these statistics potentially reflect a differential in hospitalization demand or access. In any case, they reflect horizontal inequalities in terms of health and may lead to productivity differentials between groups.

2.2. Characteristics of the working age population in terms of labor market exclusion

This section presents the characteristics of the working age population. Labor market participation is represented as a categorical variable that can take four values: 0 if a person is a non-participant in the labor market, 1 if a person is a marginal part-time worker (less than 240 hours per year), 2 if a person is a regular-part-time worker (between 240 and 1,500 hours per year) and 3 if the person is a full-time worker (1,500 hours or more)[29]. These categories respectively reflect complete labor market exclusion (category 0), partial labor market exclusion (categories 2 and 3), and inclusion in the labor market (category 4).

The variables used in the following empirical analysis are: **Gender**; **Religion and caste group** (Hindu Upper Caste, Hindu OBC, SCST, Muslim Upper Caste, Muslim OBC, Other Groups); **Highest completed education level** (None, Kindergarten, Primary, Secondary, Higher Education); **Literacy** (dummy variable indicating whether a person is literate) ; **English** (non-speaker, beginner, speaker) ; **Short-term illness** (number of days a person was sick in the last month : zero, less than one week, between one and two weeks, more than two weeks); **Long-term hospitalization** (number of months a person was in the hospital in the last year: zero, less than one month, more than one month); **Age**.

[29] The choice of the 240 hours threshold is the one chosen by the IHDS to distinguish individuals who are considered as employed or not. In the following chapters concerning labor market activity, these individuals will not be included as they were not attributed the employment questionnaire. The 1,500 hours, which is equivalent to 28.8 hours a week fall within the interval defined by the ILO for "*substantial part-time*" employment which is [21, 34].

Table 1.5. Labor market participation 2011-2012 (Row percentages)

	Non-participant	Marginal Part-time worker (]0; 240[hours)	Regular Part-time worker ([240; 1500[hours)	Full-time worker (>1500 hours)	N
Gender					
Male	17.1	4.6	10.8	67.6	22,179.0
Female	76	2.7	9.1	12.2	23,151.0
Caste/Religion Groups					
Hindu Upper Caste	50.3	3.6	7.6	38.5	12,345.0
Hindu OBC	45.4	3.9	11.2	39.5	13,872.0
SCST	42.7	3.5	11.5	42.2	9,642.0
Muslim Upper Caste	49.6	4.7	7.9	37.8	2,965.0
Muslim OBC	51.5	2.8	10.4	35.3	4,344.0
Other Groups	48.4	3.5	8.6	39.5	1,683.0
Completed Education Level					
None	57.9	3.8	13.3	25	9,004.0
Kindergarten	46.4	3.4	13.7	36.6	3,135.0
Primary	49.3	3.6	12	35.2	3,405.0
Middle	44.3	3.9	10	41.9	11,294.0
Secondary	46.4	3.7	7.8	42	6,848.0
Higher	41.5	3.4	6.8	48.3	11,580.0
Literacy					
Illiterate	57.6	3.6	13.5	25.2	9,076.0
Literate	44.5	3.6	9	42.9	36,208.0
English ability					
Non-speaker	50.1	3.7	11.4	34.8	27,996.0
Beginner	37	3.2	6.4	53.3	12,573.0
Speaker	44.2	3.7	7.9	44.1	4,713.0
Short-term illness					
Not ill	46.2	3.7	9.8	40.4	39,691.0
Ill for less than one week	53	3.5	10.5	33	4,185.0
Ill for less than two weeks	58.6	3.3	12.1	26	812
Ill for more than two weeks	54.7	3.1	12.9	29.3	642
Long-term hospitalization					
Never hospitalized	59.3	4	8.7	28	5,684.0
Hospitalized for less than one month	58.5	4.1	10	27.3	1,584.0
Hospitalized for more than one month	63.2	4.4	10.3	22.1	68

Table 1.5 continued on the next page

Table 1.5 (Continued)					
Age group					
age15_19	56.4	3	12.8	27.8	1,804.0
age20_24	52.4	2.2	9.8	35.6	4,834.0
age25_29	45.2	2.6	9.2	43	5,774.0
age30_34	40.6	3.6	9.4	46.4	5,109.0
age35_39	36.3	3.6	10.9	49.3	5,067.0
age40_44	36.9	4.1	10.9	48.1	4,678.0
age45_49	38.3	4.5	9.9	47.4	4,533.0
age50_54	37.2	4.8	10.3	47.7	3,519.0
age55_59	44.5	4.5	11.6	39.4	2,969.0
age60_64	63.6	4.2	10.6	21.6	2,451.0
age65plus	79.5	3.9	6.5	10.1	4,592.0
Total	47.1	3.6	9.9	39.3	

Source: Authors calculations of IHDS data (2011-12).
Note: Survey weights were used to provide nationally representative statistics. Row percentages mean that the sum of the percentages is equal to 100% for each row. Individuals who are enrolled in secondary of higher education are excluded from the sample.

Table 1.5 shows the row percentages for the different variables of interest by labor market status. 17.1% of men do not participate in the labor market, which is very high, considering that in 2012-13 the level of unemployment was at 2.3% according to the ILO (2016a). This discrepancy indicates a substantial inactivity rate. 76% of women do not participate in the labor market and only half of active women are full-time workers.

Concerning socio-religious groups, non-participation in the labor market is higher for Muslims OBC and Hindu Upper Castes. Full-time employment is higher for SCSTs (42.2% of SCST individuals are active in the labor market). Moreover, among uneducated individuals, 57.9% are excluded from the labor market. The share of individuals with higher education to engage in full-time work (48.3%) is higher than for any other educational attainment level. Nevertheless, an important part of this group is also unemployed. Qualified unemployment is an important issue in India as it concerns approximately 5 million graduate individuals (Unni 2016). Furthermore, inactivity is also very high in this group, especially among women who acquire education without an intention of entering the labor market. Compared to all other education levels, individuals with higher education are less often unemployed. Illiterate individuals are mostly unemployed and 50% of individuals who do not speak English are unemployed. In terms of health, labor market exclusion seems to affect individuals who were ill more than individuals who were not.

Box 1.2: Forms of unemployment and underemployment in Ranipet

The leather industry provides a large share of employment in Ranipet. As a part of a Special Economic Zone, employment generation is supposed to directly come from the industry and from its spillovers to the other sectors of the economy (Alkon 2018). Nevertheless, among the individuals interviewed, several forms of unemployment, time-related underemployment and inactivity were observed.

A factory worker we met described her vocation change after facing unemployment. She comes from a lower middle-class family and her household's investment in her education was considered an important cost. She attended nursing school directly after her secondary education, but she was unable to find a job in her sector of activity. After approximately a year of unemployment, her relatives encouraged her to look for a job in a shoe factory in her neighborhood. Through the help of other women working in the factory, she obtained an interview relatively easily and was hired immediately. Her actual salary is lower than what she would have gotten had she been a nurse. However, she claims being satisfied with her career change and would not advise any other woman to study nursing. She does not plan to ever look for a nursing job again. This experience of qualified unemployment is an example of skill mismatch which is a significant issue in India (Unni 2016), especially for women (Fletcher, Pande, and Moore 2018). In terms of human capital, not only is this woman overqualified for her actual job, but she is also in an occupation that does not require any of the skills she acquired. Furthermore, the fact that this woman does not plan to ever work as a nurse again shows a strong path-dependency to her actual occupation.

Inactivity mostly concerns women who choose to stay at home. Often these women have worked in the city's factories in the past and chose to leave their job after marriage or to take care of children. The totality of women who previously worked wish to return to a factory job when their child (or children) reach a given age. However, getting back to their previous factory job is difficult as they face important competition from younger women. All of these women think that their experience will not be considered in the determination of their salary even if they return to a factory in which they initially worked.

During the fieldwork, we did not encounter any working-age inactive or unemployed men. We did, however, interview casual workers who did not find work for the day. These men are either shoe factory workers or tannery workers, and they wait for a "*company bus*" to pick them up every morning. They do not have any information on whether the firms will require their services or not beforehand. Depending on the demand, they can remain without a job for several days. This type of work refers to time-related underemployment. Indeed, these workers are available to work more hours and they are unsatisfied with their actual working hours.

Source: Author

3. Identifying the determinants of labor market exclusion

3.1. Empirical framework

In order to better understand the possible determinants of labor market exclusion, we implement a multinomial logistic estimation. This allows us to identify the categories of the population that are significantly prone to labor market exclusion. Although we do not use a causality framework, this empirical approach sheds light on the various ways in which belonging to a gender, religion and caste group are directly and indirectly associated with labor market status. These group variables can indeed be considered as potential determinants of labor market outcome.

We estimate a multinomial logistic model where the dependent variable is labor market participation (***partlm***), a categorical variable that can take three values: 0 if a person is a non-participant in the labor market, 1 if a person is part-time worker (less than 1500 hours per year), 2 if a person is a full-time worker (1,500 hours or more).[30] The independent variables (vector X_k) are the potential determinants of labor market participation as well as a number of control variables and γ is the logistic function.

$$Pr(partlm = 1|X_k) = \gamma(\beta_0 + X_k\beta_k) \quad \text{[Eq. 1.1]}$$

Results from this estimation are interpreted using relative risk ratios that can be calculated by exponentiating the coefficient of interest: e^{β_k}.

3.2. Results

The estimation results (Table 1.6) show which variables are positively and negatively associated with each labor market status as well as the extent of these associations. Three estimations were conducted with different samples. The estimation for the sample as a whole

[30] Note that the variable of labor market participation with four categories used for the descriptive analysis, although interesting, is not adequate for a multinomial logit estimation because the Independence of Irrelevant Alternative hypothesis is not verified with the category “marginal part-time worker” being close the categories “*non-participation*” and “*part-time work*”.

is presented in Appendix 1.6. The estimations by gender group, which provides results that are clearer to interpret, are shown in Table 1.6. Since our interest in this section is to observe the channels of labor market exclusion we chose the reference category to be full-time participation in the labor market.

Table 1.6. Multinomial logit estimation results by gender

	Labor Market Participation (Reference group: Full-time worker)			
	Female Sample		**Male Sample**	
	Non-participant	**Part-time worker ([240; 1500[hours)**	**Non-participant**	**Part-time worker ([240; 1500[hours)**
Hindu OBC	-0.185***	0.248**	-0.089	0.112*
	(0.059)	(0.118)	(0.071)	(0.062)
SCST	-0.465***	-0.060	-0.108	-0.072
	(0.081)	(0.112)	(0.067)	(0.055)
Muslim upper caste	0.220	-0.047	0.150	-0.210**
	(0.141)	(0.187)	(0.098)	(0.099)
Muslim OBC	0.804***	0.586***	0.082	-0.011
	(0.126)	(0.123)	(0.116)	(0.110)
Other groups	-0.296**	-0.046	0.106	0.127
	(0.130)	(0.211)	(0.122)	(0.165)
Age	-0.257***	-0.108***	-0.259***	-0.124***
	(0.014)	(0.016)	(0.012)	(0.009)
Age squared	0.003***	0.001***	0.004***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Primary school	-0.158	0.093	-0.228	0.025
	(0.153)	(0.203)	(0.243)	(0.123)
Middle school	0.466**	0.436**	-0.309	-0.030
	(0.178)	(0.188)	(0.257)	(0.173)
Secondary school	0.424***	0.113	-0.292	-0.189
	(0.141)	(0.137)	(0.225)	(0.165)
Higher education	0.571***	-0.135	0.053	-0.236
	(0.137)	(0.227)	(0.214)	(0.174)
Literacy	0.007	-0.749***	0.197	-0.286**
	(0.138)	(0.167)	(0.233)	(0.141)
English Beginner	0.189	-0.120	-0.060	-0.039
	(0.136)	(0.158)	(0.201)	(0.175)
English Speaker	0.035	-0.002	0.103	-0.028
	(0.073)	(0.113)	(0.061)	(0.059)
Ill for less than one week	-0.847***	-0.643***	0.157	-0.306***
	(0.096)	(0.132)	(0.112)	(0.070)
Ill for less than two weeks	-0.125	0.023	-0.187***	-0.170*
	(0.075)	(0.075)	(0.067)	(0.096)
Ill for more than two weeks	0.432***	0.633***	0.405**	0.108
	(0.150)	(0.192)	(0.189)	(0.225)

Table 1.6 continued on the next page

Table 1.6 (continued)				
Married	0.035	0.162	-0.237	0.198
	(0.163)	(0.144)	(0.225)	(0.194)
Number of female children (<15 years)	1.194***	0.618***	-1.349***	-0.236***
	(0.068)	(0.083)	(0.098)	(0.075)
Number of male children (<15 years)	0.060***	0.075**	0.026	0.097***
	(0.021)	(0.036)	(0.032)	(0.027)
Constant	5.408***	1.582***	2.819***	0.642***
	(0.348)	(0.360)	(0.292)	(0.138)
State control	Yes		Yes	
Observations	22,884		21,897	
LR-Chi²	2898.76***		5856.97***	

Source: Author's calculations from IHDS (2011-12)
Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Survey Weights applied; Hausman tests were conducted for both estimations in order to test the Independence of Irrelevant alternatives between a full model and partial model. In both cases IIA holds. Note that as predicted, this assumption does not hold if we include marginal part-time work as a category.

Marginal effect computations show different scenarios and profiles of individuals, thus allowing to establish who is more likely to be excluded from the labor market.[31]

The results in Appendix 1.6. show that if all the other factors are held constant, compared to men, women are 44 times more likely to not participate in the labor market than to participate as a full-time worker. This huge coefficient is due to the fact that in the sample 86% of non-participants are women and only 16% of full-time workers are women. They are also 4.3 times more likely to participate as a part-time worker than to fully participate compared to men, all other factors being held constant. These results suggest that complete labor market exclusion is substantially determined by gender.

An increase in education is associated with an increase in the probability of not working for women, which confirms the suggestions from the descriptive statistics. Education being correlated to socioeconomic status, especially for women, those who are educated do not *need* to work to ensure a livelihood. They might stay unemployed longer in order to find a job that matches their education level. Some women are also likely to stay inactive, especially to take care of their children. It is possible that women acquire an education to find a better suitor and

[31] Relative risk ratios used to interpret these results are obtained by exponentiating the multinomial logit coefficients.

not to find a job, which increases the positive association between education level and non-participation.[32]

Results in Table 1.6 show different trends concerning socio-religious groups depending on the sample. In the female sample, being a Hindu OBC or an SCST is significantly and negatively correlated to Non-participation compared to Full-time work. However, being a Hindu OBC is positively correlated to Part-time work compared to full-time work. Furthermore, among women, being a Muslim OBC is significantly and positively associated with non-participation and part-time work in comparison to full-time work. The coefficient is however much higher for non-participation (0.804 with a 1% level significance) than for the other category. Note that these coefficients are in reference to the Hindu Upper Caste category. These results seem to suggest that female work is driven by necessity for SCSTs and Hindu OBCs, assuming that these groups suffer from lower socio-economic status. However, it is not the case for Muslim OBCs. In the male sample, Hindu OBCs are significantly (at the 10% level) more likely to engage in part-time work than in full-time work by 11.8%, compared to Hindu Upper Castes. By contrast, Muslim Upper Castes are less likely to engage in regular part-time work than in full-time work by approximately 19%. Apart from these results, there is no clear association between religion and caste groups and labor market exclusion. Indeed, none of the coefficients for the non-participant category are significant among men.

Educational attainment variables are not significant in the male subsample, except for literacy. Literate men are less likely to engage in part-time work than in full-time work when all other factors are held constant. Furthermore, short-term health-related factors indicate that men who were ill for less one and two weeks are less likely to be in full-time jobs than in part-time jobs. This result shows that health-related variables are not necessarily a predictor of productivity, health is also possibly an outcome of work.

In the female subsample, the results indicate that educated women are more likely to be non-participants in the labor market than to have a full-time occupation. Klasen and Pieters (2015) study the stagnation of FLMP from 1987 and 2011 and find that the link between rising female education and labor market participation is a combination of different mechanisms. On the supply side, the increase in the number of female graduates has increased FLMP but stigma of working in low-skilled jobs and a declining selection effect into education have attenuated this

[32] Studies that analyze the situation in which women are educated to find a suitor analyze the returns to education on the marriage market. See for instance (Maertens 2013).

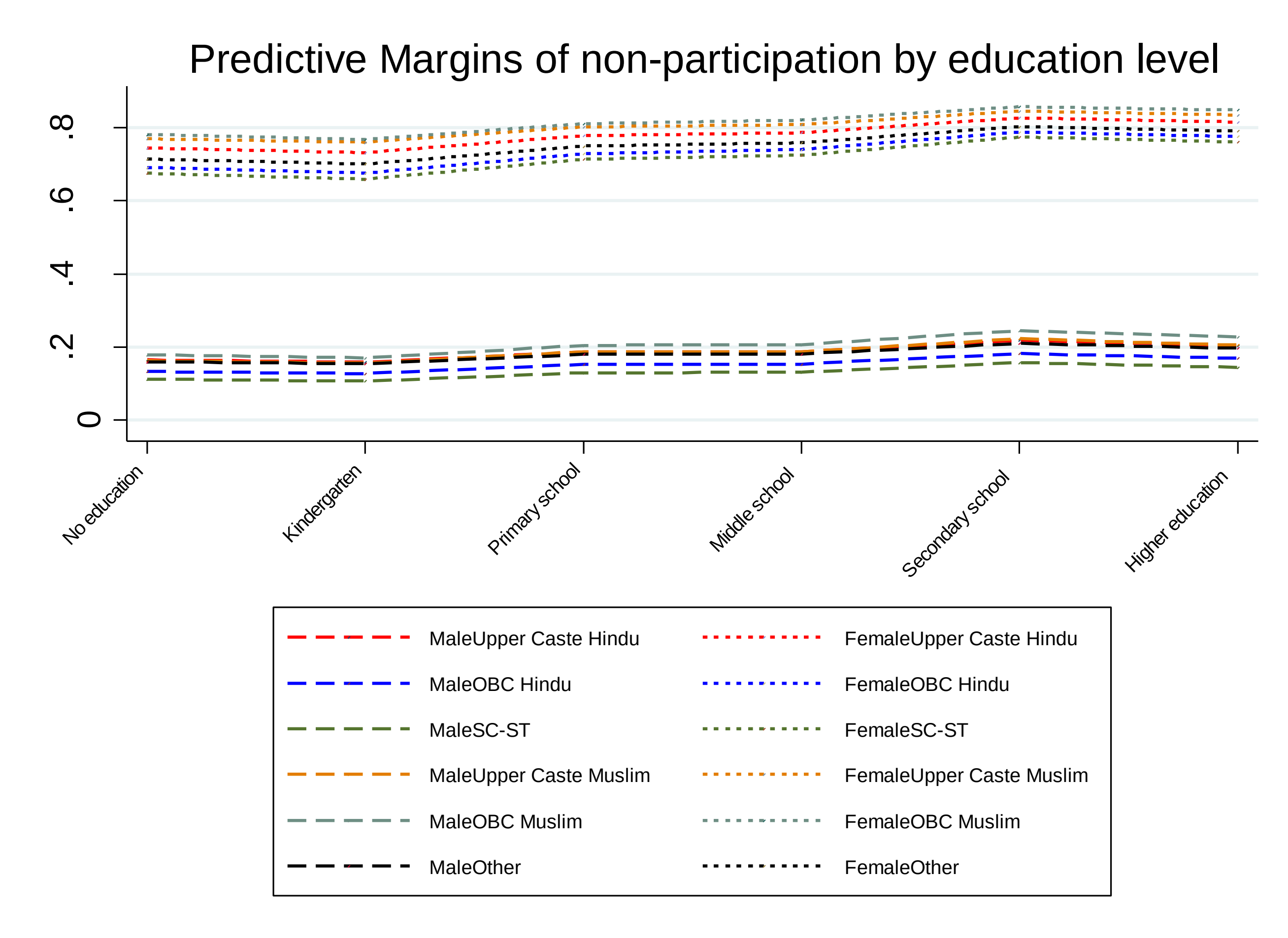
increase. On the demand side, the evolution of the sectoral structure of employment has not generated sufficient occupations for women.

In order to observe the indirect channel through which gender and socio-religious groups are associated with labor market outcomes, we compute the marginal effects[33] of non-participation for specific education and age levels. These marginal effects are calculated on the basis of the multinomial logistic estimation of the whole sample (Appendix 1.6). They show the probability of non-participation for each group at specific education levels and age values, when all other factors are held constant.

The results concerning education, presented in Figure 1.3, show that for men and women, the probability of labor market exclusion only slightly varies by the level of education. For all education levels, the difference in the probability of labor market exclusion between men and women is substantial. Caste and religion have a relatively smaller role in this case. Another striking result is the fact that the hierarchy of socio-religious groups is almost the same across all five education levels. In other words, regardless of the level of education, the probability of being a non-participant in the labor market is higher for female Muslims OBC and lower for Female SCSTs. The same trends are visible for men. An explanation for the slight increase in the probabilities of being unemployed for the higher education levels (more visible for women) is that individuals who have higher levels of education come from more economically advantaged households and they have less necessity to work.

[33] The marginal effects are obtained using the *margins* command in Stata 14.

Figure 1.3. Predictive margins of non-participation by years of education



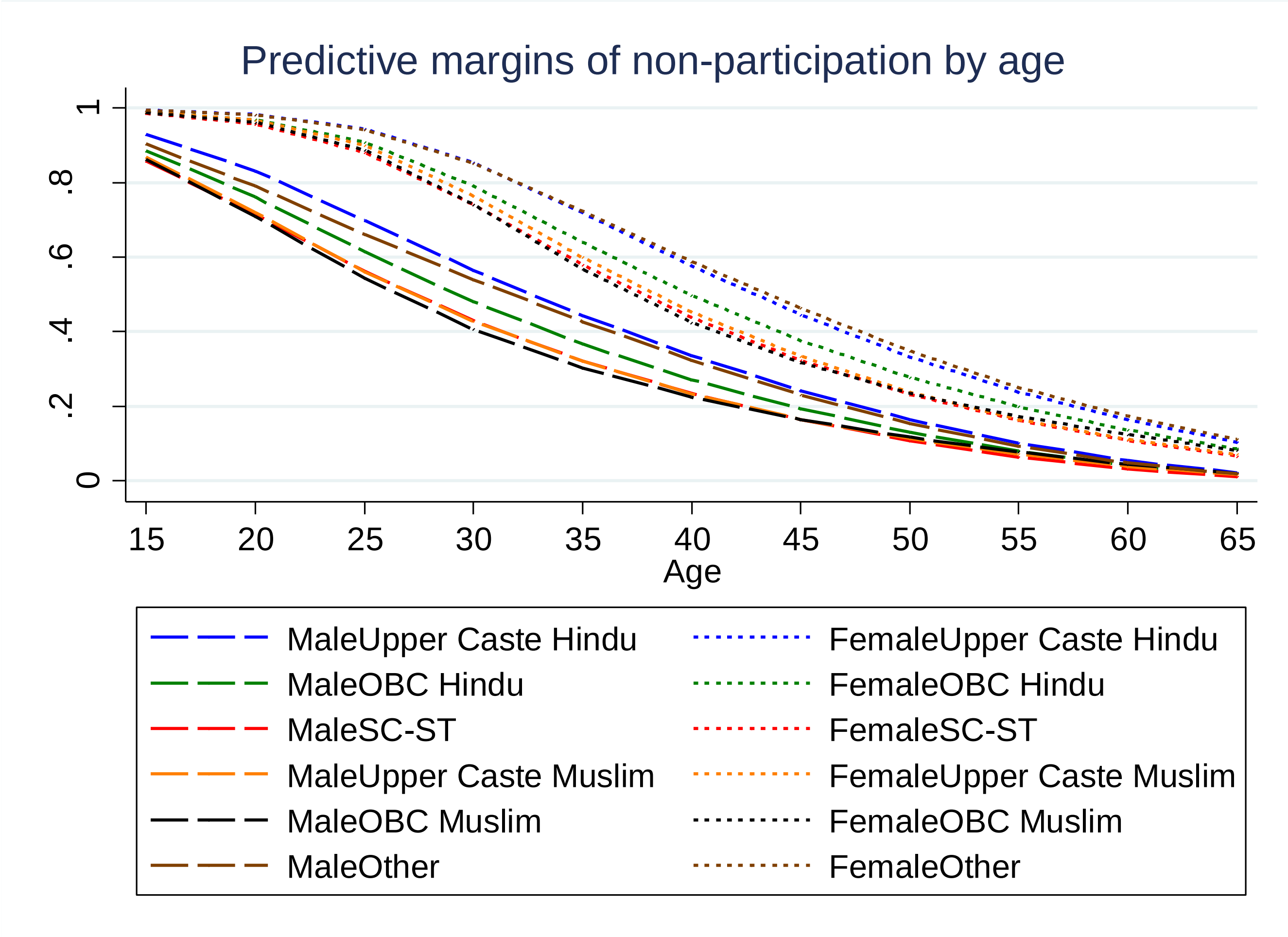
Source: Author's calculations from IHDS data

We also estimate the marginal effects of non-participation by gender, caste and religion across the age distribution in Figure 1.4. The focus on age is necessary to understand the mechanisms of labor market exclusion in India. The first interesting result is that there are two clear trends, one followed by men and one by women. The trend for women is generally concave which leads to a convergence with the male trend at the top of the age distribution. However, there is important intra-group heterogeneity in both cases. In the female group, Hindu Upper Caste women and Other Caste women have a higher probability of being excluded from the labor market all over the age distribution. This result confirms previous literature findings that considers that female labor market participation decreases for women when they are in higher caste groups (Chen and Drèze 1992), which is the case all along the age distribution. Muslim OBCs and Upper Caste Muslims have a lower probability of labor market exclusion all along the distribution, regardless of gender.

For all groups, the probability of being excluded from the labor market is substantial at younger ages. Note that the estimation from which these marginal effects are computed does not include individuals who are currently enrolled in an educational institution. These results show the importance of the youth unemployment issue raised by the Indian government. The falling participation rate of young adults is partly driven by an increase in full-time education. Inactivity has also increased among young people. However, in developing countries a large share of youth is still out of education and jobs, risking a being trapped into poverty (ILO 2016c).[34] India has the largest youth population in the world and the difficulties to include them in sustainable and decent occupations has led authors to qualify the situation as a demographic disaster rather than a demographic dividend (Mitra and Verick 2013). The economic cost of youth unemployment and underemployment is combined to non-economic costs such as crime, mental health issues, violence, drug-taking, social exclusions. Discouragement is also an issue in the case of India. Youth especially faces unequal opportunities concerning the labor market. The lack of labor market opportunities is an incentive not to invest in education (Mitra and Verick 2013).

[34] The global youth unemployment rate reached a peak in 2013 at 13.3% and despite few years of improvement, it has increased again in 2016 to reach 13.1%. Moreover, young adults are overrepresented among the unemployed as this group represents only 15% of the global labor force and 21% of the working-age population but they account for 35% of the unemployed (ILO 2016c). It is a matter of concern especially in developing countries where young people are more likely to be unemployed or to have a precarious job in the form of informal contracts (Mitra and Verick 2013)

Figure 1.4. Predictive margins of complete labor market exclusion by age



Source: Author's calculations from IHDS data

4. Discussion

This study aimed to shed light on the question of labor market exclusion. We find that gender is a strong determinant of labor market exclusion. Religion and caste are also associated with labor market status in the female sample, regardless of whether the effect is direct or indirect. In the male sample, the direct associations are quite rare and education level does not significantly change the probability of taking part or being excluded from the labor market. Other factors on the demand side may influence the probability of working not only for women, as suggested by Klasen and Pieters (2015), but also for men.

The results we provide depict the direct and indirect patterns of group differentials in non-participation. However, our empirical analysis does not establish causality between the group variables and labor market participation because of several sources of endogeneity. Regarding gender, it would be very likelihood for someone to decide on a gender change because of labor market status, remains quite low. However, the change of religion is probable and it may be driven by socioeconomic status. The most important issue in terms of causality analysis is the existence of unobserved heterogeneity. Not accounting for the motivation of soft skills, for instance, can potentially under- or over-estimate the coefficients. If disadvantaged group differs from the other groups regarding a factor that we do not account for (e.g. motivation), it only reflects an additional channel of labor market exclusion. One possible way of doing a causality analysis is to conduct an experiment such as the one implemented by Banerjee et al. (2009) in which false CVs are sent to employers and call back rates are calculated. However, even these types of studies remain inadequate when most of the labor market is informal and CVs are probably never used in the recruitment process.

Furthermore, the complexity and diversity of female labor, often taking the form of unpaid household labor in developing countries, is difficult to grasp in empirical studies. The 19th ICLS introduced a "*revolutionary*" approach to labor according to Nussbaum (Mark et al. 2017, pp. 7). Work is defined as comprising "*any activity performed by persons of any sex and age to produce goods or to provide services for use by others or for own use*." Despite the attractiveness of this definition, the lack of detailed information on household labor is the reason we use of a narrower definition of labor. Although this leads to analyses that underestimate the female labor force, policy recommendations can still be provided keeping in mind that compared to domestic work, non-domestic work is more likely to be correlated with empowerment (Fletcher, Pande, and Moore 2018).

Individuals who are excluded from the labor market consequently represent a very heterogeneous group. We explore one dimension of this heterogeneity in a study concerning the consequences of female labor market exclusion on children's education in the following section.

Section 2. Wasted potential? The gender-specific consequences of women's labor market status

The previous section has pointed out the associations between labor market exclusion and religion, caste and gender groups. This section takes a broader perspective by analyzing the possible consequences of women's labor market exclusion on their children's education.

A low FMLP has immediate effects on a country's economic development. Female labor represents an additional income for households and can contribute to GDP. Nevertheless, its intergenerational implications should also be questioned. Fernandez, Fogli, and Olivetti (2004) show that in the United States, there is a positive correlation between a woman and her mother-in-law's labor market status, suggesting that a man is more likely to choose a wife who has a larger probability of working if his own mother worked. Morrill & Morrill (2013) nuance these findings by showing that although there is a significant link between a mother-in-law's and a woman's probability of working, a woman also forms her preferences before choosing a spouse and these are influenced by her own mother's labor market experience. These findings point out the possibility for preferences to be altered by the participation of one's mother on the labor market.

Transposing such findings to developing countries, especially ones with strong gender inequality, requires caution. There is no doubt that the way gender is perceived in a society can have economic and labor-market-related consequences. In the case of India, several arguments indicate a strong bias towards women such as a preference for sons, excess female child mortality rates (Drèze and Sen 1999) or violence against women within or outside the household (Jayachandran 2015; Chakraborty et al. 2018). With respect to the labor market, this type of social norm can have consequences on productivity-related factors (e.g. education levels) and on women's decisions to work (Goldin 1995; Mammen and Paxson 2000). For instance, Chakraborty et al. (2018) analyze how sexual violence influences Indian women's decision to work. Using a cost analysis framework, by considering the risk of harassment as an additional cost, they show that women are less likely to work away from their home when the perceived threat of sexual harassment is higher.

Gender attitudes are so deeply ingrained in societies that they seem inalterable in the short term. However, an important way to promote public policy that could bring change to gender attitudes towards the labor market is to consider the way this type of norm is transmitted across generations. Finding the channels and dynamics of the transmission of gender attitudes may provide valuable insights into the way labor market policies and educational policies should be implemented. This section proposes to explore how a mother's labor market status has consequences on her children's potential labor market outcomes. The ideal way to study this effect would be to observe the labor market outcomes of children. However, given the lack of nationally representative long-run longitudinal data in India, it is only possible to observe the immediate effect on children in order to establish hypotheses about their future labor market status. Studies show that in developing countries a mother's labor market status can have ambiguous consequences on her children (Afridi, Mukhopadhyay, and Sahoo 2016; Francavilla, Giannelli, and Grilli 2013). First, female work can have an immediate positive effect on children's well-being through increased household income. It can also have a motivational effect on girls. In the long run, FLMP can increase the bargaining power of women in household decision-making and allow them to have a greater voice in education-related decisions, which has been shown to benefit children. However, women's labor takes time away from home which can have repercussions in terms of child care, the distribution of household chores among family members and child labor. In the long run, fertility decisions may also be altered because of FLMP.

The consequences of a mother's labor market status on her children's education also provide information on the transmission of gender attitudes. If women's labor market participation has an effect on education in itself, and if this effect is gender-specific, it can increase or diminish the short-term gender gap in education. Moreover, there is a longer-term perspective to keep in mind, education being not only a considerable determinant of future employability and income, but also a tool for empowerment.

Linking gender attitudes to the gender gap in education provides interesting information on household dynamics in India. All the more so given the heterogeneous nature of women's labor market exclusion that does not necessarily reflect women's own decision but can also be the consequence of a coercive household decision motivated by the negative reputation effect of female labor. The cultural practice of seclusion in some Indian households restricts women's freedom of choice and imply low-levels of decision-making power (Jayachandran 2015). This type of exclusion is a significant breach to achieving gender justice since, according to the ILO (2010), it requires that men and women be equally able to choose whether and to what extent they want to participate in the labor market.

In this study, we analyze the intergenerational consequences of women's labor market status. First, our study provides evidence on the consequences of mothers' work on children's school-related time use (school hours, homework hours and days of absence) and the general level of reading, writing and mathematics. Moreover, we observe whether these results are gender-specific. Finally, we distinguish the "*voluntary*" from the "*coercive*" non-participation of women. Note that this study uses an original question from the gender module of the IHDS database that relates whether a woman is allowed to work or not. Overall, we discuss the channels through which gender attitudes toward the labor market might be transmitted to younger generations through educational outcomes. To our knowledge, three previous studies provide information close to our research question. Kambhampati (2009) explores the correlation between women's autonomy, measured by the level of education and employment status, and children's education. She finds mixed results: first, compared to fathers, an increase in mothers' wages is associated with a higher probability of schooling but also with a higher probability of child labor, especially for girls. Francavilla, Giannelli, and Grilli (2013) study the effects of female employment on children's school enrollment. They find that the additional income provided by women's employment is usually insufficient to release children from child labor. Afridi, Mukhopadhyay, and Sahoo (2016a) observe a positive effect of women's participation in the NREGS program on time spent in school. Given the ambiguous findings on the relationship between mothers' work and children's education, the first contribution of our study is to provide additional evidence on this problematic by focusing on the gender-specific similarities and differences regarding the educational consequences of mothers' labor market participation. This analysis uses a broad variety of educational indicators reflecting children's academic level and attendance, as well as a more precise differentiation of women's labor market status. The second contribution of our study is to discuss how a woman's actual labor market status can play a role in the transmission of gender-related social norms to future generations.

Since many endogeneity issues arise from this nonexperimental setup, we use an Inverse Probability Weighting Regression Adjustment (IPWRA) method which has a "*doubly robust*" property and allows us to provide consistent estimates(Wooldridge 2010). First, we specify a treatment model where the treatment status refers to women's labor market status. Three treatment levels are distinguished: working, not working but *allowed* to work, not working and *not allowed* to work. Next, we specify an outcome model where the dependent variable reflects either education-related time use or the student's abilities. Inverse probability weights computed from the treatment model are used to estimate the outcome for each treatment level, which are used to measure the average treatment effects. The study shows that girls' test scores are negatively affected by their mother's labor market participation, whereas boys' scores are not or positively affected by it. Attendance does not seem to be particularly affected by the mother's labor market status, except for the hours of homework which are lower for children whose mothers work full-time. Overall the results suggest that girls are penalized in all cases. On the one hand, if they are in a household where female work is not stigmatized, they have lower scores. On the other hand, if they are in a household where female work is stigmatized, they probably will not be allowed to work in the future either, hence the idea of a lost potential.

1. From mothers' labor market participation to children's education: what are the transmission channels?

This section presents a brief literature review on the societal and intra-household dynamics that may constitute transmission channels from mothers' labor market participation to children's education.

1.1. Intrahousehold allocation of resources and the gender gap in education in India

The intrahousehold allocation of resources shows patterns of bias in favor of boys in patriarchal societies, especially in contexts of patrilocality[35] (Ebenstein 2014), in the following ways. Expenses regarding nutrition (Behrman and Deolalikar 1990) and the time spent for breastfeeding differ by the child's gender (Jayachandran and Kuziemko 2011). Furthermore, Asfaw, Lamanna, and Klasen (2010) provide evidence of gender discrimination in health

[35] Patrilocality can be defined as a "cultural norm in which sons provide care for their elderly parents, and daughters leave the home following marriage to provide care for their in-laws" (Ebenstein 2014, pp.3).

expenditures concerning children, which is more intense for poor households. Boys are more likely to be hospitalized than girls and parents are willing to take more risks (e.g. borrowing or selling assets to cover health costs) for their sons' health than for their daughters'. Closer to our research question, the cultural preferences for sons and the expectations that they will provide care for their parents (Ebenstein 2014) motivate households to allocate more resources for their son's education than for their daughter's (Kaul 2018).

Despite the strong political will to provide education for all children since the independence and many public policy initiatives since then, the quality of education is a cause for concern in India (Halim et al. 2016). Between 2002 and 2005, the number of out-of-school children went from 25 million to 13.5 million. However, girls seem to be particularly affected by exclusion from school. The attitude of parents and the society towards female education, children's motivation, the opportunity cost of not working or not being available for household chores are demand factors that can explain this gender inequality (Drèze and Sen 1999; Kingdon and Unni 2001). In terms of supply, girls suffer from direct or indirect discrimination. One example of indirect discrimination towards them is linked to policies for the inclusion of SCST children in schools. In more inclusive zones, where the attendance of SCST children is higher, parents from other castes tend to choose private schools for their sons, so that they do not share classes with SCST children. However, this decision is often restricted to boys because of a preference for sons in the allocation of a limited budget, which leads girls to stay in public schools (Wu et al. 2007). Assuming that the quality of education is better in private schools which have more resources than public schools, choosing different schools for boys and girls can lead to future labor market inequality because of differentials in innate ability.

The gender gap in education has many implications. It is most likely to translate into occupational and/or income differentials when children enter the labor market. Although returns to education in India are largely debated in the literature, a gender gap in education is always associated with worse labor outcomes for women than for men. Moreover, education not only provides better potential access to the labor market[36], but it also has many other roles. It contributes to subjective well-being, financial literacy, health decisions, child care or simply empowerment and is a well-established necessity for all members of society. Specific forms of education should also be considered when it comes to future labor market opportunities. One

[36] The results in Section 1 of this chapter suggest that this assumption is conditioned by the motivations of individuals relative to whether they wish to enter the labor market or not, especially for women.

interesting example is the positive correlation between mathematics test scores and future earnings (Altonji and Blank 1999; Bharadwaj et al. 2016). These scores are not only an indicator of cognitive ability, but basic calculus knowledge is also an important tool for self-employed individuals or small businesses (e.g. for account-keeping or calculating interest rates for a loan).

1.2. Consequences of female labor market participation

Many factors need to be considered when analyzing women's work-related behaviors in developing countries. The standard opposition of work and leisure merely reflects the complexity of choices women have to make. They have to decide how to allocate their time between "*market and non-market activities*" (Mincer and Polachek 1974, pp. 76), namely child care, household chores, family business work and other forms of work (Ponthieux and Meurs 2015). In all cases, children's education is likely to be affected by these decisions.

If mothers are not active in the labor market they are more likely to be present at home and have more time for childcare, which can positively impact children's education level. Moreover, when mothers do not work, they are more available for domestic chores which can benefit children. Afridi, Mukhopadhyay, and Sahoo (2016) show that, in India, educated women increase their reservation wage and choose to stay in the household as an investment for their children's education. On the other hand, the additional income provided by women's work is likely to have a positive effect on children's wellbeing which can translate into better school achievement. Nevertheless, their absence from the household may cause household chores to be transferred to children.

Francavilla, Giannelli, and Grilli (2013) find that there is a negative correlation between a mother's employment and her children's schooling. In their study, schooling is measured by a binary variable opposing children who attend school from those who do not. In the poorer households, the additional income brought by a mother's employment is not sufficient to cover the costs of schooling. The authors conclude that targeted employment policies towards women can lead to an undesirable increase in child labor and a decrease in school enrollment. The fact that children do housework allows adult members of the family to work (Cigno and Rosati 2005). Afridi, Mukhopadhyay, and Sahoo (2016) find positive effects of mothers' employment by using the variation brought by NREGS[37] take-up by rural women. In their study, FLMP is

[37] The National Rural Employment Generation Scheme (NREGS) is an employment scheme dating from 2005 consisting in the provision of 100 days of guaranteed wage employment for all working-age individuals.

associated with increased time spent in school and better grade progression. They also find that women's decision-making power increases with labor market participation, which explains the positive results. In an experimental study in three villages in rural India, Jensen (2012) finds that when presented with new labor market opportunities, women are less likely to get married and have children and more likely to study longer or enter training programs. This study shows that although the behavior of women in the household is affected by social norms, new opportunities can be a factor of behavioral change. Kalsi (2013) shows that an increase in female leadership at the local political level contributes to changing beliefs regarding sex selection. If such a motivational effect exists at the household level, female labor can have a positive effect on girls. However, this assumption relies on the empowering nature of work and supposes that it is not exclusively subsistence-related.

1.3. Identity, work and the transmission of gender attitudes

In the psychological literature, gender identity (as described in the social structural theory)[38] is the process that causes people to occupy specific roles in the society because of individual choice, sociocultural pressures or biological potentials, leading them to develop psychological qualities and in turn behaviors to fit these roles (Eagly and Wood 1999; Katz-Wise and Hyde 2010). The way gender is perceived in a society can have important implications for the labor market. Indeed, the perception of gender affects labor market participation and labor which will in turn contribute to changing the perception of gender. Gender identity particularly affects the labor market through access to occupations (information, use of social networks, relevant education level for the job market, etc.) discrimination (or self-discrimination) and segregation (or self-segregation).

The way gender identity is conceived and the labor market status can be linked through different channels. Identity affects decisions in many ways as shown by Akerlof and Kranton (2000) in their seminal study using a conceptual framework borrowed from psychology and sociology. They demonstrate that belonging to a specific group (e.g. gender or ethnicity) can lead someone to follow the behavioral pattern that is considered as normal or else this person may face an identity disutility and exclusion from others. In sociological studies, Covarrubias (2013) finds that in Mexico, social norms influence women's willingness to participate in the labor market through the internalization of moral arguments and the use of these moral arguments as a

[38] Alternative theories, namely the evolutionary theory have different definitions of the concept. For a discussion on these concepts see (Eagly and Wood 1999).

bargaining power inside of the household. Mammen and Paxson (2000) describe the different mechanisms through which social norms motivate labor market exclusion of married women in developing countries. On the one hand, women may dislike factory jobs because it is difficult and precarious. Marriage and children, therefore, become an escape from having to do these jobs. On the other hand, societies stigmatize the husbands of women who have blue-collar jobs. As pointed out by Goldin (1995), female labor shows a form of insufficiency of men's work to provide for their family. In a more recent study, the author extends the anthropological concept of the "*polluting*" nature of female work in male-dominated occupations. Women are discriminated against in occupations that are male-dominated because women doing men's jobs is perceived as a "*downgrading*" of the occupation (Goldin 2014).

In some countries, practices such as seclusion more directly prohibit female labor. Female seclusion can be defined as a practice where women are confined to the company of other women and close male relatives. In practice, this confinement can take many forms such as a veil or not being allowed to go outside of the household. According to Miller (1982), most communities in India (region, caste, class, or religious group) have their own subsystem of female seclusion, varying in the form and degree of female segregation. The most common forms of seclusion in India are *Gunghat* (in Hindu households) and *Purdah* (in Muslim households). Overall, seclusion is more strictly enforced "*in the North than in the South, among upper castes and classes than among lower castes and classes, and among Muslims than among Hindus*" (Miller 1982, pp. 780). The link between seclusion and the labor market is interesting in contemporary India. Indeed, a strict seclusion practice would imply that women do not work outside of the household. However, Miller shows that the strictness of seclusion is dependent on the degree of necessity of female labor. Although her study dates back to the 1980s, more recent analyzes point to the contemporary relevance of this social reality in India (Dhar, Jain, and Jayachandran 2015).

The transmission of gender attitudes is a long-run process involving the core institutions of a society such as religion (Psacharopoulos and Tzannatos 1989). However, studies show that in the short-run, the role of parents is key to shaping their children's gender attitudes. Dhar, Jain, and Jayachandran (2015) show that India, mothers have a greater influence than fathers on their children's discriminatory behavior. They also provide suggestive evidence that the effect is stronger on daughters than on sons.

2. Methodology

This study seeks to identify the gender-specific consequences of a mother's labor market status on the educational outcomes of her children. These outcomes are measured by a test score in mathematics, reading and writing as well as education-related time allocation: school hours per week, homework hours per week and days of absence per month. The labor market status of mothers is our variable of interest (i.e. the treatment variable). The originality of this study is to differentiate women who have access to the labor market from women who are prohibited from working because of seclusion practices and/or fear of bad reputation. This distinction implies interesting hypotheses and questions regarding potential pathways through which a mother's labor market participation can affect children's education. Among women who do not work, the difference between those who are allowed to and those who are not can provide insights on different channels that influence education outcomes: intra-household time allocation, gender-specific household investment in education and children's motivation.

Since we use observational data, many challenges arise in the identification of a potential causal effect between mothers' labor market status on children's education. In order to provide robust estimators, we use the method of Inverse Probability Weighting Regression Adjustment (IPWRA). In this section, we will present the potential bias of a simple OLS estimation, and present the IPWRA estimation method.

2.1. Baseline specification and potential bias

The linear specification that would allow identifying the effect of a mother's labor market status if individuals were randomly distributed across the categories of labor market status is the following:

$$Educ\_var = \beta_0 + \beta_1 work_j + \beta_2 X_2 + \varepsilon \qquad \text{[Eq. 1.2]}$$

$Educ\_var$ is an educational outcome of a child, $work_j$ represents the different possible work statuses of the mother and X_2 represents a set of control variables.

The dependent variables only take non-negative integer values, a Poisson estimator or quasi-maximum likelihood estimator would be more relevant in this case. Moreover, the linear estimation of mothers' labor market status on children's education suffers from an endogeneity bias for the following reasons.

The first type of endogeneity bias that raises concerns is potential reverse causality. The variable of interest is a mother's labor market status and we attempt to measure its potential effects on several dependent variables related to her children's education. The dependent variables we wish to estimate may have a causal effect on the variable of interest. The likelihood of children's test scores causing women to have a specific labor market outcome is small. Nevertheless, as shown by Afridi, Mukhopadhyay, and Sahoo (2016), some mothers choose to remain inactive as an investment in their children's education. They might be more inclined to do so if their children have good test scores. For the school attendance outcomes, the issue of reverse causality seems more likely. Whether a child goes to school or not may alter the opportunity cost of enforcing seclusion norms in households. Therefore, the probability of female labor exclusion to encourage child care can potentially be higher. On the contrary, if a child does not go to school, it may encourage the family to authorize a woman to work because the child (most probably a girl in this case) is available for household chores. The hours spent in school and the number of days of absence potentially have a similar but probably smaller causal effect. Households that need someone to do chores would most likely discourage a child from doing homework with a smaller incidence on the fact that mothers are allowed to work or not.

Besides, there is a probable selection bias both on observable and unobservable characteristics. First, unobservable variable bias is a potential issue in a linear estimation because women who work, those who do not work but are allowed and, those who do not work and are not allowed to potentially have different characteristics. This setup is prone to an omitted variable bias that we can call the *perception of gender identity* which is different, at least, in households where women are not allowed to work and households where they are. Second, there is an issue of selection on observables: this bias is caused by the fact that women from each group have different observable characteristics.

Because individuals are not randomly distributed across categories of the variable of interest, we implement a specific methodology to identify potential causal effects. In order to establish this effect, the outcome means should be unconditional of the treatment levels. In an experimental setup, the random distribution across treatment levels ensures that the treatment is independent of the outcome which makes detecting a causal effect possible, even though the counterfactual is not observed for each individual. With observational data, we do not observe the counterfactuals for each individual either (e.g. we cannot observe the score of the same child for different treatment levels) but there is also a non-random distribution of individuals across treatment levels. For this reason, we need to correct the non-randomness of the treatment assignment which will allow us to obtain estimates from an "*as good as random*" model. Inverse Probability Weighting Regression Adjustment (IPWRA) is a method that allows to do so by modeling both the outcome and the treatment. The "*resampling*" that is done through the IPWRA ensures that the data is drawn from an "*artificially random*" sample.

2.2. Inverse Probability Weighting Regression Adjustment

In cases of nonrandom assignment of a given treatment, robust estimators need to be able to measure the *unobserved potential outcomes*. To create a realistic counterfactual, a propensity score method is usually used in non-experimental data in order to identify causal effects. However, this method is limited to a treatment variable with only two treatment levels (i.e. a binary variable opposing treated to non-treated individuals). One solution to identify treatment effects when there are more than two treatment levels is to implement an inverse probability regression adjustment. IPWRA is an estimation method that corrects the endogeneity bias by modeling treatment assignment and outcomes. This method is appealing of its flexibility concerning the nature of the outcome variable since it provides the possibility of linear estimations as well as Poisson estimations.

The IPWRA estimation requires a treatment model and an outcome model. The treatment model is estimated by a multinomial logit model.

$$work_j = \gamma(\alpha_0 + \alpha_{kj} X_T + \mu_j) \quad \text{[Eq. 1.3]}$$

Where $work_{mj}$ is the multinomial treatment variable indicating the mother's labor market status, that takes j values. *j=1* if an individual is not allowed to work, *j=2* if an individual is allowed to work but does not work and *j=3* if an individual is active on the labor market. X_T are independent variables that predict the treatment: X_1 is a vector of explanatory variables, X_2 refers to control variables, α_0 the constant and ε the error term. The estimation from this treatment model is then used to derive the propensity scores p for each possible treatment level (the 3 categories of $work_j$).

$$Educ\_var = \beta_0 + \beta_k X_O + \mu \quad \text{[Eq. 1.4]}$$

The outcome model (Eq. 1.4) is estimated using a regression adjustment method.[39] $Educ\_var$ is the child's education level and the independent variables are X_O . Inverse probabilities p are

[39] A Poisson estimation can used at this step.

used as weights to correct the coefficients of the regression adjustment. X_T and X_O can overlap. The Average Treatment Effect (ATE) is calculated for each treatment level as the difference in the weighted Potential Outcome Means (POM).

The doubly-robust property of the IPWRA estimator, which stems from modeling both treatment and outcome models, implies that only one of the two previous equations (1.3 and 1.4) need to be correctly specified for the procedure to yield consistent estimates (Cattaneo 2010; Wooldridge 2010). In our case, the treatment model is the one with the less potential endogeneity, since many of the variables we use are calculated at the PSU level[40]. Hence, they are less likely to be affected by the dependent variables in the treatment model. Moreover, the ignorability assumption must be verified. This means that the treatment is assigned at random, conditional on a set of observable characteristics. We use a rich set of covariates to verify this assumption at its best. Balance tests are also conducted to detect the presence of a selection effect into different treatment levels.

3. Model specification and descriptive analysis

We use the 2011-12 wave of the IHDS database.[41] The "*Women*" module of the IHDS questionnaire provides data on issues pertaining to fertility, marriage and gender relations in the household. This information was obtained by questioning a random ever-married woman aged 15 to 49 in the household. To make sure that we only capture an intergenerational effect, only women from age 15 upwards are considered as mothers, and the dependent variables concern children below 15 years old. Since we use control variables from the IHDS I (2005 wave) database in one set of estimations, the sample of children is restricted to individuals between 6 and 15 years.

---

[40] PSU indicates the Primary Sampling Unit used in the collection of the IHDS data. It is the intermediate level between the household and the district.

[41] Note that this is the only study in which we include rural areas to ensure that we have sufficient observations.

3.1. Variable description for treatment and outcome models

Table 1.7 presents the variables used to estimate the treatment and outcome models.

Table 1.7: Variables used for the treatment and outcome models

Treatment model (the mother)	**Outcome models (the child)**
Treatment status	**Outcome indicators**
Mother_Work1: 1. Not allowed to work; 2. Allowed to work and not working; 3. Working	Model 1. *General score:* Total test score (reading, writing, mathematics) [0;9]
Mother_Work2: 1. Not allowed to work; 2. Allowed to work and not working; 3. Working part-time[42]; 4. Working full-time	Model 2. *Mathematics score:* Mathematics score only [0;3]
	Model 3. *School Hours:* Hours spent in school/week
	Model 4. *Homework Hours:* Hours spent doing homework/week
	Model 5. *Days of absence:* Days of absence/month
Covariates	
Age: Mother's age	*Age*: Child's age
Educ_GM1: Maternal grandmother's education	*School_level*: Highest Completed school level
Educ_GM2: Paternal grandmother's education	*N_sisters*: Number of sisters (<15 years old)
HHmale_educ: Highest Male education in the household	*N_brothers*: Number of brothers (<15 years old)
Rel_caste: Religion and Caste group	*School_dist*: School distance in Km
PCCons2005: Per capita consumption (in 2005)	*N_ill*: Number of days of sickness in the last 30 days
Poor2005: Poor household (in 2005)	*HHmale_educ:* Highest adult male education in hh
Coverhead: "*ghungat*"/"*burkha*"/"*purdah*" or "*pallu*" (PSU level)	*HHfemale_educ:* Highest adult female education in hh
Land_dec: Mother involved in land-related decisions (PSU level)	*HHhead_emp*: Employment situation of household_head
Health_dec: Mother involved in health-related decisions (PSU level)	*Rel_caste*: Religion and Caste group
Husb_violence1 and *Husb_violence2*: Husband violence (PSU level)	*PCCons2005*: Per capita consumption (in 2005)
Region	*Poor2005*: Poor household (in 2005)
Urban	*Region*
	Urban

Source: Author's calculations from IHDS data

[42] The definition of part-time work is the same as in Section 1. A part-time worker has less than 1500 hours of work per year.

The originality of our study lies in the variable used to analyze labor market exclusion which is a combination of being allowed to work and employment status. The variable **Mother_work1** is a categorical variable that indicates if a mother is "*Not allowed to work*," "*Allowed to work and not working*" or "*Working*." We test an additional specification with another treatment variable **Mother_work2** which additionally differentiates women who work part-time from those who work full-time. Concerning the educational outcomes, we consider gender-specific outcomes variables related to school attendance and general level in reading, writing and mathematics. To measure the general level of education, we calculate a **general score** as the sum of the scores in mathematics, reading and writing and we also measure the specific effect on **mathematics score**. To measure school attendance, we use variables indicating the number of hours spent in school during a week (**school hours**), the number of days of absence per month (**homework hours**), and the number of hours spent doing homework (**days of absence**).

The outcome variables are expressed in a count data format (hours or score) which is why we use an IPWRA Poisson estimator.

3.2. Model specifications

The treatment and outcome models have confounding variables as well as specific variables.

In the treatment model, the independent variables are a vector of determinants of FLMP and a set of control variables. **Educ_GM1** and **Educ_GM2** (the education level of the mother's mother and mother-in-law respectively) are included following the findings from Morrill and Morrill (2013) and Fernandez, Fogli, and Olivetti (2004) which indicate the positive correlation between a woman and her mother's and mother-in-law's employment status. Since we do not have the information about the employment status of the woman's relatives, we add their education level as a proxy. We also introduce another potential determinant of FLMP: the education level of the husband (because of possible assortative matching) or of the father in the household: **Educ_male (**highest male education in the household). We chose not to add the employment status of the household head in order to limit the possible endogeneity in the relationship between this variable and a woman's labor market status. Indeed, Miller (1982) shows that in rural India, the strictness of seclusion practices can be dependent on how much a woman's work is necessary. We also include **Religion_caste**, a multinomial variable indicating religion and caste groups, to control for specific behaviors in the different socio-religious groups. Moreover, the variable **Coverhead** is a dummy variable indicating the percentage of households in each PSU in which women have to cover their head. This practice which is also called *ghungat*, *burkha*, *purdah* or *pallu* is closely related to seclusion. The following control variables, also calculated at the PSU level, indicate women's bargaining power and autonomy in households: **Dec_land** and **Dec_health** (indicate whether a husband takes decisions regarding land investments and health center visits); **Husb_violence1** and **Husb_violence2** (respectively indicate whether a woman thinks it is normal for a husband to beat his wife if she leaves the house without permission and if she does not cook properly). Two variables controlling for the economic status are added **PCcons2005** (average per capita consumption in 2005); **Poor2005 (**dummy variable indicating whether the household is below the poverty line in 2005 using the Tendulkar poverty line). Additional control variables such as the woman's age, the region and whether the household is in an urban area were also added.

In the outcome model, a gender-specific analysis is provided for each **Educ_var (score_total, score_maths, hours_school, days_abs, hours_hw)**. The independent variables for the analysis are the following: **Age; school_level** (highest completed school level); **N_sisters** and **N_brothers** (number of sisters and brothers in the household below 15 years old); **Hheduc_f** and **Hheduc_m** (highest adult female and male education in the household); **Hh_Head_work** (employment situation of the male household head); **School_dist** (distance to school in km); **Ndays_ill** (number of days the child was ill in the last month); **PCcons2005** and **Poor2005**, **Religion_caste, Zone** and **Urban**.

Child_labor, which indicates the work status of the child, is added in a separate series of estimation, as a robustness analysis. It is prone to measurement error and does not consider all forms of non-remunerative labor (e.g. household chores), which can lead to an underestimation of its effects.

3.3. Descriptive Statistics

Table 1.8 presents the distribution of mothers across the categories of *Mother_Work1*. The sample is composed of 40.81% mothers who work full time or part-time.

Table 1.8. Labor market participation of mothers

Mother_Work1	N	Percentage
Not working and not allowed to work	7091	19.00
Not working and allowed to work	15004	40.19
Working	15235	40.81
Total	37330	100.00

Source: Author's calculations from IHDS data

Table 1.9 presents the descriptive statistics for attendance levels by gender. Among children younger than 15 years old, the average number of hours spent in school in a week is 32.081 (32.099 for girls and 32.061 for boys). On average, children spend 7.8 hours doing homework per week. The highest difference in terms of attendance is visible for the days of absence, although it remains quite small in absolute terms (3.601 days for boys and 3.583 for girls). The gender gap in scores is more important (Table 1.10). Girls systematically have lower scores than boys. The difference between the average summed score for girls and boys is 0.209.

Table 1.9. School attendance rates

	School Hours (per week)		**Homework Hours (per week)**		**Days of absence (per month)**	
	N	Mean	N	Mean	N	Mean
Male	19421	32.099	19251	7.847	19082	3.601
Female	17567	32.061	17398	7.844	17295	3.564
Total	36988	32.081	36649	7.846	36377	3.583

Source: Author’s calculations from IHDS data

Table 1.10. Score in reading, mathematics and writing

	General Score ([0;9])		**Reading Score** (0- cannot read; 1-can read letters; 2- can read words; 3- can read paragraph; 4- can read a story)		**Maths Score** (0 – cannot recognize numbers; 1- recognizes numbers; 2- substractions; 3-divisions)		**Writing Score** (0- cannot write; 1-few mistakes; 2- no mistake)	
	N	Mean	N	Mean	N	Mean	N	Mean
Male	6058	5.283	6134	2.579	6108	1.562	6077	1.142
Female	5578	5.074	5658	2.505	5634	1.450	5596	1.121
Total	11636	5.183	11792	2.544	11742	1.508	11673	1.132

Source: Author’s calculations from IHDS data

As a preliminary step to our analysis, we verify whether the treatment statuses have significantly different effects on the educational outcomes[43]. Since the latter variables follow a Poisson process, we estimate each of them with one independent variable (work1) using Poisson regressions (Table 1.11). The results show that there is indeed a significant (at the 1% level) education gap between each possible outcome of the treatment variable except for one case: the general score for boys is not significantly different when they have a mother who is allowed to work but is inactive compared to when they have a mother who is not allowed to work.

[43] The motivation for this step is equivalent to conducting a Student’s T-test between treatment groups for normally distributed variables.

Table 1.11. Poisson estimation of the treatment status on educational outcomes

Variables	General score (girls)	General score (boys)	School hours (girls)	School hours (boys)	Homework hours (girls)	Homework hours (boys)	Days absence (girls)	Days absence (boys)
Work1 (reference: not working and not allowed to work)								
Not working and allowed to work	-0.122***	-0.014	0.015***	0.016***	-0.131***	-0.114***	0.148***	0.118***
	(0.019)	(0.018)	(0.005)	(0.004)	(0.009)	(0.008)	(0.015)	(0.014)
Working	-0.232***	-0.143***	0.042***	0.031***	-0.238***	-0.215***	0.177***	0.191***
	(0.019)	(0.019)	(0.005)	(0.004)	(0.009)	(0.008)	(0.015)	(0.014)
Constant	1.772***	1.732***	3.446***	3.454***	2.217***	2.200***	1.112***	1.143***
	(0.016)	(0.016)	(0.004)	(0.004)	(0.007)	(0.007)	(0.013)	(0.012)
Observations	4,068	4,526	12,700	14,082	12,583	13,964	12,478	13,821

Source: Author's calculations from IHDS data
Note: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 1.12. Descriptive statistics

	Treatment level 1 Not allowed to work		Treatment level 2 Not working but allowed to work		Treatment level 3 Working	
Number of observations	2521		4758		9046	
	Mean	SD	Mean	SD	Mean	SD
Age	10.75	2.31	10.44	2.37	10.65	2.29
Mother's Age	35.71	5.94	35.19	5.65	35.80	6.01
School years	4.71	2.48	4.23	2.56	4.47	2.43
School distance (Km)	2.17	3.30	1.90	2.62	1.93	3.41
Number of sisters	1.21	1.07	1.38	1.17	1.35	1.13
Number of brothers	1.33	0.93	1.45	1.02	1.40	0.93
Number of days ill	0.84	2.58	1.21	3.03	1.10	2.78
Household head works						
part-time	0.31	0.46	0.35	0.48	0.50	0.50
full-time	0.57	0.50	0.54	0.50	0.45	0.50
Highest male adult education in household (years)	8.14	4.74	7.64	4.78	5.66	4.70
highest female adult education in household (years)	6.08	4.87	5.55	4.94	3.43	4.40
OBC Hindu	0.31	0.46	0.31	0.46	0.37	0.48
SCST	0.19	0.40	0.25	0.43	0.42	0.49
Upper Caste Muslim	0.11	0.31	0.12	0.32	0.03	0.17
OBC Muslim	0.12	0.33	0.09	0.28	0.05	0.22
Other	0.03	0.16	0.03	0.16	0.01	0.11
Per capita household consumption (INR, 2005)	816.61	808.68	727.12	551.67	545.87	435.87
Poor (Tendulkar, 2005)	0.26	0.44	0.27	0.45	0.40	0.49
Urban	0.50	0.50	0.41	0.49	0.17	0.38

Source: Author's calculations from IHDS data

The descriptive statistics from Table 1.12 show differentials between the households depending on the treatment status. Working women are more often from poor households. Moreover, the households where women are not allowed to work have a higher per capita household consumption level.

4. Results

This section presents the estimation results. In the first set of estimations, women's labor market status is a categorical variable (*Mother_work1*) in which we separate women who are allowed to work from women who are not. In a second set of estimations, we additionally separate women who work part-time from those who work full-time (*Mother_work2*).

4.1. School level and attendance: baseline estimations

Table 1.13 presents the IPWRA estimations with independent variables from both waves of the IHDS dataset. The Average Treatment Effect (ATE) is the specific effect of the being assigned to one treatment level on the outcome. In other words, it is the effect of the mother's labor market status on the child's educational outcome. The Potential Outcome Mean (POM) the average outcome for each treatment level. The detailed estimation results are available in Appendix 1.7.

The results show that the sign of the ATE differs by gender in several cases, which is a first indication of the relevance of calculating gender-specific educational outcomes. Compared to girls whose mothers are not allowed to work (which is the reference group for the estimation), the general score of girls in the two other groups are significantly and negatively affected. Out of a total possible score of 9, the ATEs are 0.541 and 0.516 (both being significant at a 5% level) for the treatment levels "*allowed to work and not working*" and "*working*" respectively. On the contrary, boys' general score is higher when their mother is allowed to work (whether they actually work or not) with an ATE of 0.354 (significant at a 5% level) for inactive mothers and an ATE of 0.317 for active mothers (significant at a 10% level).

Table 1.13. Average treatment effects

Treatment effects (base outcome: Not allowed to work)						
			Girls			Boys
	Not allowed to work (base outcome)	Allowed to work and not working	Working	Not allowed to work (base outcome)	Allowed to work and not working	Working
			General score			
ATE	N.a.	**-0.541** (0.222)**	**-0.516** (0.219)**	N.a.	**0.354** (0.165)**	**0.317* (0.163)**
POM	5.474 (0.210)	4.993 (0.081)	4.958 (0.075)	4.940 (0.152)	5.294 (0.074)	5.257 (0.070)
			Mathematics score			
ATE	N.a.	**-0.135* (0.0739)**	**-0.152** (0.0730)**	N.a.	0.089 (0.0560)	0.070 (0.0552)
POM	1.719 (0.077)	1.557 (0.050)	1.536 (0.054)	1.684 (0.090)	1.686 (0.047)	1.659 (0.047)
			School hours			
ATE	N.a.	0.010 (0.363)	0.368 (0.365)	N.a.	-0.514 (0.371)	0.024 (0.376)
POM	32.225 (0.321)	32.236 (0.173)	32.593 (0.179)	32.843 (0.336)	32.329 (0.158)	32.866 (0.171)
			Homework hours			
ATE	N.a.	-0.036 (0.283)	-0.435 (0.272)	N.a.	-0.311 (0.260)	**-0.460* (0.256)**
POM	8.428 (0.245)	8.392 (0.146)	7.993 (0.125)	8.464 (0.322)	8.153 (0.123)	8.004 (0.115)
			Days of absence			
ATE	N.a.	-0.0400 (0.224)	-0.109 (0.219)	N.a.	0.0104 (0.213)	0.264 (0.213)
POM	3.517 (0.202)	3.477 (0.101)	3.408 (0.090)	3.377 (0.192)	3.387 (0.093)	3.640 (0.096)

Source: Author's calculations from IHDS data
Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

The estimation shows significant differentials between the scores of girls whose mothers are not allowed to work and the scores of girls whose mothers are allowed to work but do not work. Note that the results control for the level of per capita consumption and poverty status of the family. This means that this effect does not reflect an income differential between households. One possible explanation is that there are different patterns of intra-household resource allocation or differences in other non-pecuniary behaviors between the two types of households. This is probably led by the fact that for girls who come from households with seclusion there is a necessity for girls to be more educated for reputation or marriage purposes.

Following the findings from (Bharadwaj et al. 2016), we conduct our analysis specifically on the mathematics score. As for the general score levels, we find that the effect of mothers' labor market participation is negative for girls. Interestingly, when mothers are allowed to work but do not work, there is still a negative effect on girls' scores. No significant treatment effect is found for boys. The POMs indicate that when their mothers are allowed to work, the scores are higher than daughters, which is probably due to the difference in household characteristics. This indicates that a mother's permission to work and her labor market participation seems to be associated with increasing the gender gap in mathematics, with a significant negative effect on girls. Since mathematics knowledge is an indicator on the ability to succeed in the labor market (in terms of earnings for instance) mothers' labor market participation has a potentially negative effect on girls' potential to earn in the future, which is a paradoxical result.

The results concerning school attendance show that there is no significant effect of mothers' labor market status on the hours spent in school and on the days of absence. Time spent to do homework is negatively affected (at a 5% level) by mothers' labor market participation.

The detailed results (in Appendix 1.7) show that, compared to Upper caste Hindus, belonging to the SCST group negatively influences children's scores when their mothers do not work but are allowed to, or when their mothers work. For all other groups, there is no significant effect on girls' scores. Girls attendance (school hours) is negatively affected when they come from an Upper Caste Muslim group and their mothers are outside of the labor market (whether they are allowed to work or not). A similar but significantly smaller effect is visible for boys but only in the case where their mother is outside of the labor market but is allowed to work.

These results point out the relevance of a gender-specific analysis as it clarifies some of the ambiguities present in the literature. The positive effect for boys with active mothers on the labor market concurs with the findings from Afridi, Mukhopadhyay, and Sahoo (2016) and the negative effect found for girls echoes the study from Francavilla, Giannelli, and Grilli 2013).

4.2. Differentiating part-time from full-time labor market participation

Table 1.14. Average treatment effects with part-time/full-time differentiation

	Treatment effects (base outcome: Not allowed to work)							
	Girls				Boys			
	Not allowed to work	Allowed to work and not working	Working part-time	Working full-time	Not allowed to work	Allowed to work and not working	Working part-time	Working full-time
				General score				
ATE	N.a.	**-0.541****	**-0.532****	**-0.435***	N.a.	**0.356****	0.260	**0.363***
		(0.227)	**(0.235)**	**(0.242)**		**(0.166)**	(0.174)	**(0.189)**
POM	5.466	4.925	4.934	5.030	4.934	5.290	5.194	5.297
	(0.216)	(0.081)	(0.100)	(0.116)	(0.152)	(0.074)	(0.092)	(0.118)
				Mathematics score				
ATE	N.a.	**-0.137***	**-0.150***	-0.0939	N.a.	0.090	0.034	0.086
		(0.076)	**(0.078)**	(0.0812)		(0.056)	(0.059)	(0.064)
POM	1.548	1.411	1.398	1.454	1.469	1.559	1.503	1.555
	(0.0710)	(0.029)	(0.035)	(0.041)	(0.0511)	(0.026)	(0.032)	(0.040)
				School hours				
ATE	N.a.	0.0260	-0.0180	**0.790****	N.a.	-0.513	-0.0943	0.491
		(0.364)	(0.508)	**(0.402)**		(0.373)	(0.472)	(0.424)
POM	32.218	32.244	32.200	33.009	32.836	32.322	32.741	32.327
	(0.322)	(0.173)	(0.395)	(0.244)	(0.339)	(0.158)	(0.332)	(0.257)
				Homework hours				
ATE	N.a.	-0.032	-0.267	**-0.706****	N.a.	-0.307	-0.259	**-0.634****
		(0.282)	(0.346)	**(0.295)**		(0.260)	(0.282)	**(0.288)**
POM	8.414	8.382	8.147	7.707	8.453	0.146	8.194	7.818
	(0.246)	(0.143)	(0.246)	(0.167)	(0.232)	(0.122)	(0.165)	(0.174)
				Days of absence				
ATE	N.a.	-0.040	-0.095	-0.167	N.a.	0.007	0.258	0.284
		(0.224)	(0.229)	(0.261)		(0.213)	(0.225)	(0.285)
POM	3.518	3.478	3.424	3.351	3.384	3.390	3.641	3.667
	(0.201)	(0.101)	(0.111)	(0.168)	(0.192)	(0.093)	(0.119)	(0.212)

Source: Author's calculations from IHDS data
Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

In the estimation presented in Table 1.14, the treatment variable is further decomposed by differentiating mothers who work part-time from mothers who work full-time. It provides more precision concerning the fact that a mother's availability can affect children's attendance, probably through an increase in the time spent doing household chores. This could explain the only significant results regarding attendance which indicate a negative effect on homework hours for children with full-time working mothers. This effect is stronger for girls (-0.706, significant at the 5% level) than for boys (-0.634, significant at the 5% level). Note that the average time spent doing homework in the whole sample is 7.85 hours per week.

Furthermore, the negative effect on girls' general score is higher for the treatment status "*Working part-time*" than "*Working full-time*." This implies that girls whose mothers are part-time workers are more penalized in terms of scores. Replacing the mother for household chores does not seem to solely explain the negative effect since we would expect the magnitude of the effect to be higher for girls with full-time working mothers in this case. Other factors such as intra-household resource allocation or motivation can differ among the groups.

4.3. Robustness tests

This section addresses several concerns about the robustness of the treatment effects. We discuss the inclusion of child labor as a control variable. Next, we verify whether there is an attrition bias due to the use of control variables from the first wave of the IHDS data. We also present balance tests, following which we discuss a possible bias due to the fact that we conduct the estimations at the child's level and several children have the same mother.

4.3.1. Adding child labor as a control variable

The existence of child labor may influence the results. Nevertheless, information on child labor potentially suffers from measurement error since families may not clearly provide information on this subject. In the sample of children used in this study, only 3.03% work. Note that nonetheless, this statistic is not too far from the 3.9% of child labor calculated by ILO (2017) using Census 2011 data.

Table 1.15. Average treatment effects with part-time/full-time differentiation and child labor

	Treatment effects (base outcome: Not allowed to work)							
	Girls				**Boys**			
	Not allowed to work	Allowed to work and not working	Working part-time	Working full-time	Not allowed to work	Allowed to work and not working	Working part-time	Working full-time
General score								
ATE	N.a.	**-0.586*** (0.225)**	**-0.597** (0.232)**	**-0.435* (0.241)**	N.a.	**0.444*** (0.165)**	**0.342** (0.173)**	**0.504*** (0.187)**
Mathematics score								
ATE	N.a.	**-0.160** (0.0755)**	**-0.178** (0.0781)**	-0.098 (0.0816)	N.a.	**0.116** (0.0564)**	0.0565 (0.0594)	**0.134** (0.0638)**
School hours								
ATE	N.a.	0.0689 (0.359)	-0.0454 (0.501)	**0.824** (0.400)**	N.a.	-0.551 (0.372)	-0.167 (0.473)	0.535 (0.419)
Homework hours								
ATE	N.a.	-0.0847 (0.282)	-0.375 (0.347)	**-0.780*** (0.294)**	N.a.	-0.270 (0.265)	-0.262 (0.286)	**-0.518* (0.292)**
Days of absence								
ATE	N.a.	-0.0554 (0.226)	-0.0432 (0.232)	-0.164 (0.262)	N.a.	-0.0580 (0.219)	0.253 (0.235)	0.266 (0.291)

Source: Author's calculations from IHDS data
Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

In households where women are not allowed to work, it is less probable to find a working daughter. Therefore, working girls will most probably be present in the two types of household, thus explaining lower scores. However, working boys could be in all three groups. In terms of attendance, no significant effect is found in the baseline estimations except for the hours of homework for boys, which are lower for girls with working mothers. When we add child labor as a control variable, we can see that the ATEs on score are larger and the significance is also higher or similar (Table 1.5). The coefficient for girls' homework is also significant. These results are therefore robust to the inclusion of child labor.

4.3.2. Controlling for possible attrition

The results presented until now exclude children between 6 and 15 years old who come from households that were not interviewed in the first wave and for whom we do not have information about consumption expenditure and poverty status in 2005. In the case of non-random attrition, the exclusion of these children from the main sample may bias the results.

In order to do so, we verify if the dependent variables are significantly different in the whole sample and the active sample.

Table 1.16. Full sample means and Balanced sample means of educational outcomes

Variable	**Full sample mean (Std dev.)**	**Balanced sample mean (Std. Dev.)**	**T-Test p-value**
General score girls	5.075 (2.746)	5.038 (2.750)	0.497
General score boys	5.283 (2.684)	5.265 (2.701)	0.710
Mathematics score girls	1.562 (0.965)	1.555 (0.964)	0.662
Mathematics score boys	1.450 (0.963)	1.438 (0.958)	0.520
School hours girls	32.379 (8.612)	32.414 (8.445)	0.731
School hours boys	32.383 (8.534)	32.466 (8.501)	0.394
Homework hours girls	8.084 (6.059)	8.227 (6.052)	**0.053**
Homework hours boys	8.069 (6.044)	8.160 (5.994)	0.190
Days of absence girls	3.502 (4.800)	3.545 (4.860)	0.464
Days of absence boys	3.526 (4.933)	3.547 (4.920)	0.711

Source: Author's calculations from IHDS data

The results show no significant differences in the means of the dependent variables except for girls' homework hours. Their level of homework hours is smaller in the balanced sample than in the full sample. In order to explore how this significant difference may affect the results, we calculate the means of girl's homework hours by treatment status (Table 1.17). The means are significantly different in all cases except for mothers who work full-time, which is the category for which we find a significant negative effect. Overall, the comparison tests suggest that this result would be stronger if there were no attrition.

Table 1.17. Mean of girls' homework hours by treatment status

Category of mother's labor market status	Mean of full sample	Mean of balanced sample	T-test p-value
Not allowed to work and not working	9.184 (7.272)	9.767 (7.426)	**0.024**
Allowed to work but not working	8.056 (5.842)	8.513 (6.087)	**0.000**
Working part-time	6.862 (5.089)	7.342 (5.080)	**0.000**
Working full-time	7.802 (6.050)	8.085 (0.154)	0.162

Source: Author's calculations from IHDS data

4.3.3. Balance tests

Balance tests reporting the raw and weighted averages of the differences between the different treatment levels are presented in Appendix 1.8[44]. They show relatively small differences for most variables[45], especially after the reweighting, except for the variable *Poor2005*. Indeed, the girls who are from households in which women are allowed to work and do not work are poorer than the rest. This suggests that our significant results may reflect economic inequality in the households. Other differences are present for caste and religion groups and for regional control suggesting that if there is an unobserved heterogeneity bias, it is probably linked to these variables. Such a bias can be the perception of gender in the community or the region, which can affect educational outcomes as well as the mother's labor market status.

4.3.4. Estimations with first-born child

A potential concern is that the standard errors cannot be clustered at the household level to account for the fact that several children may have the same mothers. Since the mothers' labor market status is estimated in the treatment model which is used to calculate propensity scores, the outcome models are partly immune to this issue. Nevertheless, in order to verify the

[44] We only show the balance tests for general score because the tests are very similar across the different estimations

[45] The weighted average should be close to 0 and the weighted variance should be close to 1. We considered 0.15 gaps from these values to be concerning.

robustness of the results, we implement the estimations for the first-born sons and daughters only.

Table 1.18. Treatment effects on first-born child

Treatment effects (base outcome: Not allowed to work)						
	Girls			**Boys**		
	Not allowed to work (base outcome)	Allowed to work and not working	Working	Not allowed to work (base outcome)	Allowed to work and not working	Working
			General score			
ATE	N.a.	**-0.642*** (0.215)**	**-0.720*** (0.219)**	N.a.	0.079 (0.224)	0.015 (0.222)
POM	6.061 (0.181)	5.419 (0.133)	5.341 (0.142)	4.940 (0.152)	5.548 (0.127)	5.612 (0.129)
		N=822			N=902	
			Mathematics score			
ATE	N.a.	**-1.619* (0.089)**	**-0.183** (0.090)**	N.a.	0.003 (0.099)	-0.024 (0.098)
POM	1.719 (0.077)	1.557 (0.050)	1.536 (0.052)	1.684 (0.090)	1.686 (0.047)	1.659 (0.047)
		N=831			N=912	
			School hours			
ATE	N.a.	0.225 (0.570	0.880 (0.557)	N.a.	**-1.591** (0.652)**	-0.456 (0.653)
POM	31.605 (0.467)	31.830 (0.333)	32.485 (0.311)	33.342 (0.570)	31.750 (0.314)	32.885 (0.326)
		N=2,121			2,289	
			Homework hours			
ATE	N.a.	-0.078 (0.449)	-0.683 (0.428)	N.a.	-0.760* (0.461)	**-0.953** (0.451)**
POM	9.089 (0.377)	9.009 (0.255)	8.406 (0.214)	9.163 (0.404)	8.403 (0.231)	8.210 (0.214)
		N=2,106			2,280	
			Days of absence			
ATE	N.a.	0.361 (0.366)	0.119 (0.347)	N.a.	0.057 (0.362)	0.190 (0.640)
POM	3.193 (0.309)	3.555 (0.199) N=2,085	3.312 (0.164)	3.389 (0.326)	3.446 (0.159) N=2,244	3.579 (0.166)

Source: Author's calculations from IHDS data
Note: Robust standard-error in parenthesis

Table 1.18 shows that the negative impact on girls' score is higher when only the first-born daughter is concerned. Indeed, a significant (at the 1% level) negative ATE of 0.642 and 0.720 is found when mothers are allowed to work and not working and when they work, respectively. However, the results concerning the score of boys disappears and a significant negative effect for treatment level 2 ("*Allowed to work and not working*") is found on the school hours of boys.

The results from this estimation confirm that the negative effect on the general and mathematics score for girls is present and significant but it does not allow us to verify the magnitude of the coefficients from our baseline results. Note that it is not surprising to find different magnitudes when restricting the sample to first-born children. Indeed, Kaul (2018) shows that first-born sons receive preferential treatment regarding educational expenditure and school enrolment in India.

5. Discussion

The results presented in Section 2 sheds light on the way labor market exclusion of women can have gender-specific consequences on educational outcomes. The aim of the study was to analyze three possible transmission channels through which a mother's labor market status can have an effect of children's education: (1) intra-household time allocation, (2) gender-specific patterns of household investment in education and (3) children's motivation. Concerning intra-household time allocation, the results show that time spent doing homework for boys and girls is negatively affected by their mother's full-time work. This result suggests that chores are transferred to all children when mothers work full-time. The mother's labor market status does not affect on the days of absence. The results concerning the general score shed light on two possible transmission channels: a difference in the way households invest in education or a difference in children's motivation. Being allowed to work negatively affects the score of girls, regardless of whether the mothers work or not. This result indicates that girls from these households probably benefit from fewer resources than girls from other households. Since our estimations control for the level of household income (although imperfectly as shown in the balance tests), the differences may pertain to the intra-household distribution of resources. It is possible that in the households where mothers are not allowed to work, the reputation of girls is important for future marriage decisions and the marriage market, and education is a positive reputation signal. Furthermore, we expected girls who are from households in which their mother work to benefit from a positive motivation effect, through an increase in their scores. However, our results indicate the contrary. Several scenarios could explain this result. First, having a mother who is secluded possibly motivates girls to perform better at school to increase their chance of escaping seclusion themselves. Having a mother who is allowed to work can also demotivate girls from schooling because of the accumulation of "*work*" inside of the household and schoolwork. Moreover, the empowering nature of work may not be relevant in the case of India, where the majority of working women work for reasons linked to necessity, which is why no motivation effect is visible. In their study, Dhar, Jain, and Jayachandran (2015) find that girls with more gender-progressive parents intend to stay longer in school than those with gender-discriminatory parents. Since we find that girls from households in which women are not allowed to work are likely to have better scores, we can consider that they are in gender-discriminatory households and they intend to drop-out earlier than other girls. This paradoxical effect contributes to the premarket gender inequality is labor supply.

Although we attempt at correcting the endogeneity bias, unobserved heterogeneity may persist. Unobserved factors may simultaneously have a negative effect on girls' score and encourage FLMP. Apart from the observed differentials in terms of income, socio-religious identity and region in the balance tests, potential unobserved heterogeneity can be linked to the intra-household allocation of resources.

To conclude, the study in this section shows that there is an intergenerational adverse association between female labor participation and educational outcomes. The results show that girls are penalized in all cases. On the one hand, if they are in a household where work is not stigmatized, they have smaller scores. On the other hand, if they are in a household where work is stigmatized, they have higher scores. Moreover, a male bias is visible in score but not in school-related time use.

Conclusion of Chapter 1

This chapter provides a twofold analysis of labor market exclusion in India. After a conceptual review allowing to understand the place of inactivity, unemployment and time-related underemployment, we establish the profiles of individuals who are excluded from the labor market. The results indicate that being a woman is a strong determinant of labor market exclusion, regardless of education level. Gender and caste seem to interact in interesting ways as female individuals from Hindu OBC and SCST groups are less likely to remain outside of the labor market compared to Hindu Upper Castes. The contrary is visible for Muslim Upper Caste women. The interaction between gender and caste does not affect the participation of men (with the exception of Hindu OBC men who are more likely to be part-time than full-time workers and Muslim Upper Castes men for whom it is the contrary). Surprisingly, education level is not a strong indirect determinant of labor market exclusion. More universal effects are visible in terms of age. The associations between age and labor market exclusion follow the same trends for all groups although a hierarchy seems to exist in the magnitude of this probability.

In the second section, we provide an analysis of the gender-specific consequences of mothers' labor market status on her children's educational outcomes. We provide evidence of clear and robust differentials in terms of general score, girls having higher scores when their mothers are not allowed work and boys lower. This surprising effect points out probable differentials in the intrahousehold distribution of chores and incomes. Our study does not specifically isolate women who are secluded. They are either not allowed to work because of seclusion or because of a fear of a bad reputation. Further investigations could provide interesting results regarding the different mechanisms between both types of households.

In a nutshell, this chapter suggests that premarket factors have an enormous role in shaping individuals' educational level, motivation and willingness to enter the labor market. Individuals enter the labor market with considerable inter-group differences and their status in the labor market, in turn, plays a role in the intergenerational transmission of educational gaps. The following chapters will analyze if there is a persistence of horizontal inequalities inside of the labor market.

Chapter 2. An analysis of labor market mobility in urban India

1. Introduction

The study of labor market mobility falls within the scope of social mobility concerns and represents an important vector of equity. Being "*an avenue to long-term equality*," upward mobility has social benefits such as increasing motivation, a positive effect on aspirations and social cohesion. It is also economically efficient since it supposes that workers will be more productive if their talent is allocated accordingly (Rama et al. 2014, pp. 164). Upward labor market mobility holds many promises in terms of social welfare given the role of careers in shaping opportunities to move up the social ladder.

In India, studies show that high growth rates are concomitant with increasing inequality since the 1980s (Sandip Sarkar and Mehta 2010; Motiram and Sarma 2014). The combination of these two trends implies that all individuals do not benefit from better economic conditions. Observing whether these inequalities are horizontal is key to understanding the dynamics of the Indian economy. Given the persistence of unequal opportunities in the access to basic needs and education in India, social mobility is a way of achieving equality of outcomes despite initial inequality of opportunity (Rama et al. 2014). The employment structure of India remains dominated by agriculture with 58% of the Indian workforce being in this sector in 2004-05 followed by services (23%) and manufacturing (11.7%). Between 2004-05 and 2011-12, despite a high average growth rate, employment generation was relatively low (the total workforce was respectively 459.1 and 474.2 million workers) leading economists to qualify this evolution as "*jobless growth*." The consequences of the global crisis of 2008 were less severe on GDP than for other countries because of India's lower dependency on exports. However, there was a significant decline in foreign investments and in the global demand for manufacturing products, which impacted employment (OECD 2010). In 2011-12, agriculture was still predominant in the employment structure (48.9%), followed by services (26.8%) and manufacturing (12.3%) (Mehrotra et al. 2014).

Several factors indicate that the allocation of workers across occupations is highly influenced by social identity which causes workers to sort into specific occupations (on the supply side) and can also lead to behaviors of discrimination or nepotism (on the demand side), the two being influencing each other in the long-term. In the case of India, as shown in chapter 1, social norms such as seclusion also mean that many female workers are not entirely (or at all) in charge of their work-related decisions. Belonging to specific socio-religious groups can also influence occupational choice, especially since career-related path-dependency is an inherent dimension of the Indian caste system. In this context, a focus on labor market mobility in urban India can provide insights on how caste-based occupational specializations are evolving in contemporary India. Labor market mobility, especially if it benefits vulnerable groups, is a sign of more equitable socioeconomic development. Observing patterns of labor market mobility can also inform policy-makers on the efficiency of implemented policies (e.g. affirmative action policies) on labor market outcomes.

The analysis of any form of social mobility requires adopting either an intergenerational or intragenerational perspective. The first consists of measuring mobility between two generations (Jäntti and Jenkins 2015) and the latter "*refers to observed differences in the economic circumstances of individuals over time*" (Burkhauser and Couch 2012). As for the dimensions of labor market mobility, two categories of trends can be observed: occupational mobility, which can be defined as a career evolution or change in occupations (Crespo, Simoes, and Moreira 2014) and income mobility, which is due to an absolute or relative income change in time.

Occupational groups are important indicators of social stratification which is why economists and sociologists often use occupational change as a metric for the measurement of social mobility (Long and Ferrie 2013; Rama et al. 2014). Beyond their economic role, it is necessary to consider their empowering nature, their quality and how they contribute to defining identity, especially in contexts where other aspects of social identity (e.g. castes) are deeply related to occupations (Akerlof and Kranton 2000). Generally, career evolutions imply higher income, increased responsibilities and sometimes better working conditions. An industry change, on the other hand, is harder to interpret. It may be correlated to better opportunities in another career path or it may be motivated by unfavorable economic cycles as a strategy to avoid unemployment (Evans 1999).

Multiple methods can be used to analyze income mobility, reflecting different definitions of the concept such as positional change, income growth, overall inequality or income risk (Fields and Ok 1999; Jäntti and Jenkins 2015). This study proposes to focus on positional change. Although the other concepts of mobility are equally interesting, their implementation with a two-wave dataset can be inadequate. For instance, comparing income growth requires at least three periods. Positional change measures the *relative* mobility of an individual in the income distribution between (at least) two dates (Jäntti and Jenkins 2015; Brunetti and Fiaschi 2013).

Taking advantage of the panel dimension of the IHDS dataset, we propose to analyze the medium-run labor market transitions between the two waves of data (2005 and 2011-12). Besides providing a descriptive analysis of the patterns of labor market mobility in urban India, the aim of this study is to observe its determinants and the extent to which these patterns vary by gender, religion and caste. We contribute to this literature by adopting a labor market perspective which combines an analysis of occupational mobility (namely between casual and permanent occupations, industries and skill levels in occupations) and hourly earnings mobility. In comparison to measuring *intertemporal household income mobility*, the analysis of *intertemporal rank change in the distribution of hourly earnings* provides more information on how labor markets are a vector of social mobility. Moreover, focusing on workers' hourly earnings has a practical appeal as it allows to compare patterns of labor market mobility between men and women. To our knowledge, very few studies address the microeconomic dimensions of *intragenerational occupational labor market mobility* and *intragenerational relative income mobility* in urban India.

We propose a methodology that takes into consideration several econometric issues. First, we use a Heckman method to correct the selection bias that arises from non-random labor market participation and selective attrition. This study also proposes to extend the bootstrapping estimation method to control for the sensibility of earnings which are prone to measurement error, especially for hourly earnings. By generating the distributions of hourly earnings in the bootstrapping process, the relative mobility we measure is less likely to be spurious. Furthermore, in order to identify the determinants of labor market mobility, a bias linked to initial earnings needs to be corrected. This variable is a potential determinant of occupational and hourly earnings mobility. In both cases, the variable is either theoretically (simultaneity bias) and/or mechanically (used in the derivation of hourly earnings mobility) endogenous to the labor market mobility outcomes. Although the aim of this paper is not specifically to address a causality between initial earnings and labor market mobility, not dealing with this endogeneity bias would result in incorrect estimates for all of the other variables. We attempt at correcting this bias using a control function approach (in the estimation of occupational mobility) and an instrumental variable approach (in the estimation of earnings mobility). Tests regarding the identification power of the instruments indicate the robustness of the estimates.

The analysis we conduct allows to observe (1) patterns of mobility for different gender, religion and caste groups, (2) if gender, religion and caste are determinants of occupational or earnings mobility, (3) the role of other potential determinants such as previous earnings and human capital variables and (4) the correlations between occupational and earnings mobility.

2. Labor market mobility, income mobility and occupational mobility: an overview of the literature

Labor market mobility is a concept that encompasses *income mobility* and *occupational mobility*. If the concept of occupational mobility (i.e. transitions across occupations) is rather straightforward, income mobility has multiple dimensions. It is possible to analyze individual income growth, the reduction of inequality associated with the longitudinal averaging of income[46] and positional change.[47] The latter dimension of mobility can be defined as a "*pattern of exchange of individuals between positions, while abstracting from any change in the concentration of people in a particular slot in each year. The latter change is "structural mobility," whereas the former is "exchange mobility" [...]. Changes in income affect positional mobility only insofar as these changes alter each person's position relative to the position of others. Equiproportionate income growth or equal absolute additions to income for everyone raise incomes, but there is immobility in the positional sense.*" (Jäntti and Jenkins 2015, pp. 811). Most studies measure the mobility of household income and fewer focus on individual wages since their aim is often to understand socioeconomic mobility such as descends into (or ascends out of) specific groups such as poverty, vulnerable groups or the middle-class(Dang and Lanjouw 2018). To our knowledge, only a few studies analyze income mobility by focusing on individual earnings or hourly wages (see for instance Buchinsky and Hunt (1999)).

[46] There are two types of mobility linked to the reduction of inequality. The first one considers income as a fixed sum of individual-level permanent and transitory components of income. The second one considers an idiosyncratic transitory component (Jäntti and Jenkins 2015).

[47] Jäntti and Jenkins (2015) propose a detailed review of the concepts and methods relative to income mobility in chapter 10 of the Handbook of Income Distribution.

Studies of income mobility show that there are important movements across the income distribution. Dang and Lanjouw (2018) find sizeable transfers out of poverty into vulnerable groups and out of the latter group into the Indian middle-class between 2009-10 and 2011-12. Although most households seem to benefit from upward movements, the intensity varies across socio-religious groups. Azam (2016) and Ranganathan (2016) study household income mobility using IHDS data and their results are contradictory as to which strata of the population benefits more from mobility. Indeed, Azam (2016) finds that Hindu Upper Castes have the highest probability of upward mobility (and the lowest probability of downward mobility) in rural and urban areas. He also finds that SCSTs face the lowest upward mobility in rural areas and Muslims face the lowest upward mobility in urban areas. By contrast, Ranganathan finds that in rural areas, mobility is higher among backward castes.

The literature usually provides separate analyses of income and occupational mobility except for Altonji, Smith, and Vidangos (2009) who develop a model in which they simultaneously estimate earnings, employment, job changes, wages and work hours in the context of the United States between 1975 and 1996.

Monsen, Mahagaonkar, and Dienes (2012) study occupational transitions into self-employment in India using the National Data Survey on Savings Patterns of Indians (2004-05) dataset. They find that regional economic factors such as self-employment rates and unemployment rates affect the likelihood of transition from salaried employment to self-employment, higher unemployment being negatively correlated to the intention and the effective transitions into self-employment. They also find that higher self-employment rates are negatively correlated to intentions and effective transitions into self-employment. They do not explore whether these patterns differ by religion and caste groups and they also exclude women from the study because of their scarce transitions into self-employment (only three women among 3,144 individuals are concerned in their sample).

The caste system implies that jobs are determined at birth (Deshpande 2000) making occupational specialization one of its inherent characteristic. Two main trends have altered the link between caste and the labor market. Colonialism has made the concept of caste a structuring factor of the Indian society (Dirks 2001), but the modernization of the Indian economy that has been deploying since the 1980's has, on the contrary, rendered caste-based occupational specialization more flexible according to policymakers. Modernization does not only weaken barriers of entry in specific occupations, but it also creates new forms of employment. However, facing modernization, the caste system adapts and rearranges (Harriss-White 2003) to create new forms of employment segregation. In this context, detecting occupational mobility is a way to observe the ever-changing relevance of the caste system as an economic characteristic in India. Mobility has increased over time and across generations. Using IHDS (2005) data, Motiram and Singh (2012) show that there is a large occupational path-dependency across generations, specifically for low-skilled occupations. Indeed, 55.87% of farmers have fathers who held the same occupations as theirs. This percentage increases up to 62.4% for non-agricultural self-employed workers. Lanjouw, Murgai, and Stern (2013) compare data from the 1950s to 2008 to observe the economic situation of various castes in the village of Palanpur (Uttar Pradesh). They find that the historically most deprived caste (*Jatabs*) have known significant mobility, partly due to more opportunities in non-farm occupations. Moreover, father-son occupation associations for cohorts between 1945 to 1984 show that more than 40% of the sons of unskilled fathers held more qualified occupations (Motiram and Singh 2012). However, Reddy (2015) shows that in recent years (1983 to 2012) occupational mobility has declined more sharply for SCSTs.

Women are likely to be more present in temporary occupations because of interruptions linked to pregnancy and child-care. Temporary jobs are generally characterized by a higher turnover, especially when they are precarious and do not require specialized skills since changing occupations does not imply a high cost (Arulampalam and Booth 1998; Crespo, Simoes, and Moreira 2014). On the other hand, occupational segregation, taking the form of self-selection of women into specific jobs because of beliefs regarding "*male*" and "*female*" jobs (Goldin 2014) as well as possible barriers of entry (e.g. not meeting educational requirements, lack of experience, insufficient social network or discrimination) can lead to reduced mobility across occupations, but also in terms of income.

Patterns of female income mobility are less analyzed than socio-religious income mobility. In India, studies usually use the equivalized household income to measure income mobility which by definition does not take gender into account. Moreover, studies of occupational mobility compare situations of men with an intergenerational (men compared to their fathers) or intragenerational (personal trajectories) perspective. Nevertheless, women's labor market mobility can have interesting implications at the household level, but it may also attenuate the sharp contrasts between socio-religious groups. A context-specific study conducted by Luke and Munshi (2011), among tea-plantation workers in South India who benefit from permanent employment and equal wages across castes, shows that women's economic mobility contributes to reshaping the decision-making process in the household.

Box 2.1. The links between labor market mobility and geographical mobility in Ranipet

An interesting characteristic of the town of Ranipet, which is populated by 51,000 inhabitants, is that it lies between the rural and metropolitan India (Denis and Marius-Gnanou 2011). This town is a part of the "*middle India*" described by Harriss-White (2003), and one of its characteristics is its dynamism. The industrial district of Palar valley, where Ranipet is located, is an important contributor to the global leather production. More than 50% of India's leather is produced in this zone (Amelot and Kennedy 2010). Ranipet is a manufacturing center that produces finished leather and shoes for the European market. The growth of the leather industry, encouraged by an increase of Foreign Direct Investment, has increased employment opportunities in the town and has been an incentive for migration inflows. Two types of geographical mobility (i.e., migration), motivated by social and occupational mobility, can be identified.

The interviews have indeed pointed out a common form of intergenerational labor market mobility, led by rural to urban migration. Indeed, many middle-aged workers from SCST groups migrated to Ranipet from more or less geographically close rural areas in the State of Tamil Nadu. The migration took place during their youth, often before marriage, and was motivated by a will to escape poverty (i.e., push factors) and seize opportunities in the leather manufacturing industry (i.e., pull factors). In these cases, the parents of the migrants were initially agricultural workers. Furthermore, after finding a job and settling down, the wives of the married men joined their husbands in Ranipet. Some single men went back to their village to get married, following which the wife migrated to Ranipet. In this case, migration has been associated with intergenerational mobility in terms of well-being regardless of the quality of the jobs they hold in Ranipet. All workers consider that their situation is better than if they had followed their parents' path. This form of migration is therefore also linked to occupational mobility, from agricultural to manufacturing activities. The migration of a wife in order to follow her husband is also likely to be a case of occupational change or entry in the labor market. In some cases, we were able to meet the wives in question. These women were active in the labor market, and when asked if they would have had a job in their hometown, they all answered that they would be agricultural workers or inactive.

Another very different form of migration concerns workers from the State of West Bengal. These workers migrate in groups, through a contractor, and work for a few months in a factory. Concerning the interviewed workers, the period of the working arrangement was three months. The working agreement of these workers can be renewed in the same factory. Otherwise, they can find another similar occupation in another factory in Tamil Nadu or another State. In this type of careers, a worker's geographical mobility is not necessarily associated with social mobility or mobility in terms of the skill requirement of the occupation. Indeed, they often have to execute the same type of tasks from one factory to another. Nevertheless, the fact that these workers sometimes learn to speak Tamil (the main language in Tamil Nadu) is a form of skill development. The potential use of this skill in the labor market remains however limited because factories hire bilingual individuals as middle-men to manage this type of migrant groups.

Source: Author

3. Analyzing mobility with the IHDS dataset

Using the IHDS dataset, the aim of our study is to analyze medium-run labor market mobility in urban India. This study uses the two waves (2005 and 2011-12) of the IHDS database which has a panel data structure. In order to analyze the urban labor market, the analysis focuses on individuals who live in urban areas in both waves. The sample is restricted to individuals who were at least 15 years old in 2005. To analyze occupational mobility, we only include salaried workers, whether they receive a regular wage or have a casual form of employment. We also use information on earnings exclusively from salaried employment since business and self-employment earnings are calculated at the household level and we do not have the information at the individual level. Moreover, we exclude the bottom and top percentiles of the wage distributions in order to exclude extreme values. Based on the different variables used in the analysis, the final count of individuals for whom we have information in both waves can be reduced up to 6,471, depending on the specification. The remainder of this section presents the variables that will be used to detect occupational and earnings mobility and assesses the existence of selective panel attrition.

3.1. Occupational variables

We use the following information from the IHDS database to identify casual and regular forms of labor, sector of activity (hereafter referred to as "*industry*") and occupation type in both waves:

- Classification of occupations between *casual labor* (which includes hourly and daily wage employment) and *regular labor* (which includes regular, long-term or permanent employment arrangements).
- The two-digit identification of the *industry* of a worker from the National Classification of Industries (NCI 1987). From the initial 97 groups of occupations, we construct five groups: Agriculture, Manufacturing, Services, Public Administration and Construction.
- The two-digit identification of individuals' *type of occupation* from the National classification of occupations (NCO 1968). The National Classification of Occupations of 1968 has been modified several times which has the advantage of making it more practical. The NCO of 2004 has an interesting feature as it allows to classify occupations by skill level. We recoded the two-digit 1968 NCO to the 2004 NCO[48]. It allows classifying individuals into nine occupational groups (presented in Table 2.1), each referring to a skill requirement level. This level can take the values 1 (low skill requirement) to 4 (high skill requirement).

Table 2.1 shows the distribution of individuals across the categories of these variables for 2005 and 2011-12.

The analysis of earnings mobility is conducted using the hourly earnings variable provided by the IHDS data for 2005 and 2011-12.[49] Table 2.3 presents the statistics of earnings for both years.

[48]We used the code equivalency table established by D'Agostino (https://anthonylouisdagostino.com/resources-code/). Since the IHDS database only contains the 2-digit classification, there are some ambiguities for a few occupational groups for which had to be split into 2 different groups depending the 4-digit classification of 2004. In these cases, we used the verbal declaration of occupation as an additional information and we recoded the 1968 2-digit into the 2004 2-digit group that reflected the most frequently declared activity. These ambiguities do not affect the broader 4-skills classification obtained from the NCO (2004), which is used for this analysis.

[49] Note that this variable is consistent with other earnings variables in the dataset. Indeed, it is equal to the yearly wage divided by the total of hours worked.

Table 2.1. Occupation variables

Group	Skill Level	N (Balanced panel sample)	Percent (2005)	Percent (2011-2012)
Industry		**6471**		
Agriculture			8.69	6.20
Manufacturing			18.14	22.86
Services			53.09	49.50
Public Administration			5.89	7.79
Construction			14.18	13.66
Total			100	100
Occupation type (NCO 2004)	**Skill Level**	**6930**		
Legislators, Senior Officials and Managers	4		2.87	6.00
Professionals	4			
Associate Professionals	3		12.40	16.08
Clerks	2			
Service Workers and Shop & Market Sales Workers	2			
Skilled Agricultural and Fishery Workers	2		60.64	62.18
Craft and Related Trades Workers	2			
Plant and Machine Operators and Assemblers	2			
Elementary Occupations	1		24.07	15.74
Casual-Permanent		**7027**	**Percent (2005)**	**Percent (2011-2012)**
Casual worker			69.10	49.82
Permanent worker			30.89	50.18

Source: Author's calculations from the IHDS dataset

3.2. Panel data description and attrition

An issue pertaining to the use of unbalanced panel data is the existence of attrition. The second wage of the IHDS data contains 83% of the households from the first wave. If the attrition is randomly distributed across the variables of interest, it does not affect the results. Conversely, the existence of *non-random* or *selective attrition* can cause the estimates to be biased. For instance, individuals who migrate and cannot be found in the second wave may not be randomly distributed across the categories of occupational change and across the distribution of hourly earnings mobility. Moreover, since individuals who do not work are not included in this study, transitions out of the labor market are part of the attrition issue.

As suggested in Nguyen, Nordman, & Roubaud (2013), it is possible to verify the existence of selective attrition by comparing the means and distributions of the relevant variables from the whole samples of each wave and the subsamples that will constitute the balanced panel.

Table 2.2 shows that there are significant differences in the distribution of individuals across the occupational mobility categorical variables between the full and balanced panel samples. We compare the shares of individuals in each category of the occupation variables. Two-sample Z-tests[50] show significant differences in the shares of permanent workers, skill levels 1 to 3, agriculture and construction workers for 2005. They also show significant differences for skill levels 1 to 3, agriculture, services and public administration workers for 2011-12.

Table 2.2. Comparison of full and balanced samples

Variable	Categories	Percent in full sample (2005)	Percent in balanced sample (2005)	P-value of proportion tests
Casual-permanent	Casual	68.57	69.11	0.248
	Permanent	31.42	30.89	**0.000**
Occupation by skill level	Skill level 1	25.08	24.07	**0.055**
	Skill level 2	59.13	60.64	**0.010**
	Skill level 3	12.78	12.40	**0.367**
	Skill level 4	3.01	2.87	0.504
Industry	Agriculture	9.38	8.69	**0.057**
	Manufacturing	18.40	18.14	0.607
	Services	53.67	53.09	0.355
	Public Administration	5.69	5.89	0.501
	Construction	12.85	14.18	**0.001**
		Percent in full sample (2011-12)	**Percent in balanced sample (2011-12)**	**P-value of proportion tests**
Casual-permanent	Casual	50.34	49.82	0.468
	Permanent	49.66	50.18	0.468
Occupation by skill level	Skill level 1	17.17	15.74	**0.001**
	Skill level 2	59.66	62.18	**0.000**
	Skill level 3	17.19	16.08	**0.040**
	Skill level 4	5.98	6.00	0.947
Industry	Agriculture	5.23	6.20	**0.004**
	Manufacturing	22.88	22.86	0.958
	Services	53.24	49.49	**0.000**
	Public Administration	6.21	7.78	**0.000**
	Construction	12.43	13.66	0.013

Source: Author's calculations from IHDS data

Concerning hourly earnings, Student T-tests (Table 2.3) conducted between the full samples and balanced samples show that there is no significant difference in the hourly earnings of 2005 but there are significant differences for 2011-12. Selective attrition is therefore suspected since those who are in the balanced sample earn more than those who are not. However, kernel density plots in Figure 2.1 show that the attrition does not affect the distribution of hourly

[50] We conducted z-tests of two shares for each category using the *prtest* command in Stata.

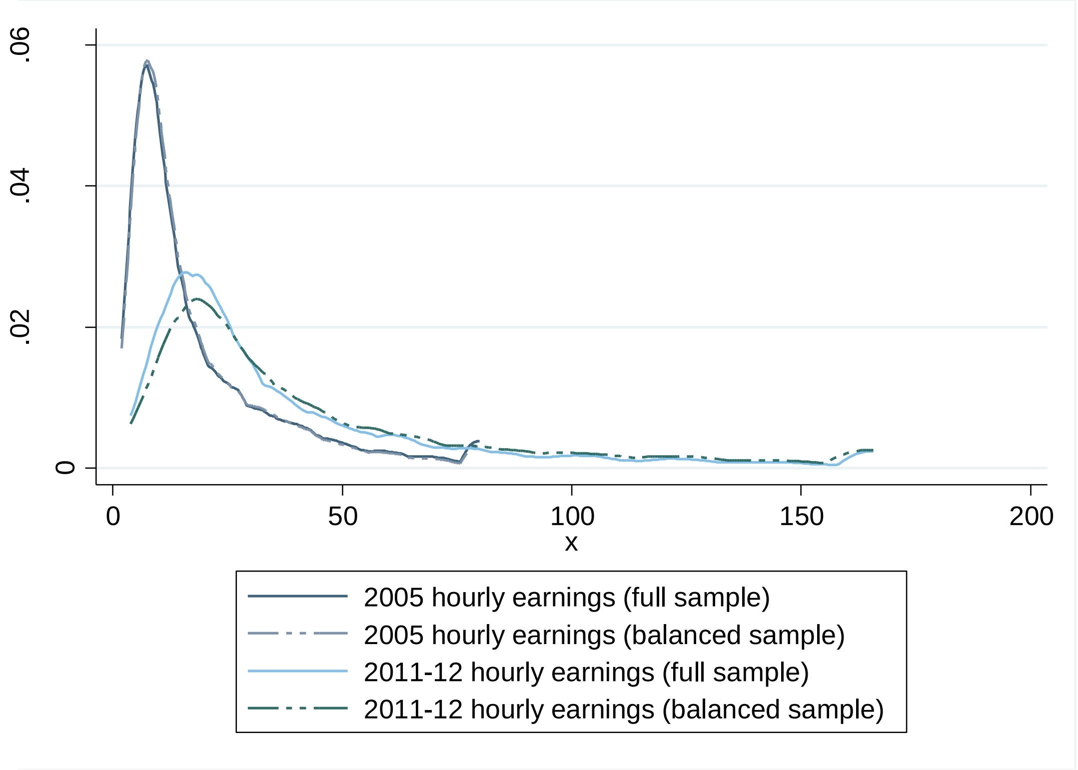
earnings for 2005 but the earnings of 2011-12 have a slightly larger right-tail in the balanced sample.

Table 2.3. Hourly earnings in full and balanced samples

Variable	N	Mean	Sd	Difference
Hourly earnings full sample 2005	9812	19.01	17.43	0.23
Hourly earnings balanced sample 2005	6992	18.78	16.74	
Hourly earnings full sample 2011-12	11513	39.51	36.18	4.50***
Hourly earnings balanced sample 2011-12	6620	44.01	39.18	

Source: Author's calculations from IHDS data

Figure 2.1. Distribution of hourly earnings between full sample and balanced sample in both waves



Source: Author's calculations from IHDS data

The investigation presented in this section point to the existence of selective attrition, to which we will propose a solution involving a Heckman selection correction method.

4. Methodology for analyzing medium-run labor market mobility

Several methods can be used to detect labor market mobility. The methodological choices allowing its detection and measurement can lead to significantly different interpretations (Fields 2006). Labor market mobility can be measured in *absolute* terms (e.g. movements across a poverty threshold) or in *relative* terms, the latter being more adapted to numerical outcomes than to categorical ones. We choose to adopt an *absolute approach* to measure *occupational mobility*, allowing us to observe transitions in and out of different types of occupations. Conversely, to study *income mobility,* we choose to adopt a *relative approach* in order to observe relative movements of workers across the earnings distribution. These two approaches provide complementary information that we further analyze in a third step by estimating the correlations between both types of mobility.

4.1. The detection of labor market mobility

Occupational mobility refers to job-related transitions between two dates. Using an absolute approach, our aim is to distinguish individuals who “*moved from*” and those who “*stayed in*” a given group between two dates. For this purpose, as presented in the previous section, we establish a classification of occupations and industries and distinguish casual workers from permanent ones.

For each category, the method to detect any movement is to establish transition matrices between the two dates. These matrices allow computing the row percentages of movers and stayers. In the case of a two-wave dataset (t=1; t=2) with two professional statuses A and B, Table 2.4 shows that the workers who kept the same status in both periods are the *stayers* and that workers who changed statuses are the *movers*. If we assume that the professional status can be ranked, B being better than A, it is possible to distinguish an *upward mobile* group (status in t=2>status in t=1) from a *downward mobile* one (status in t=1>status in t=2).

Table 2.4. Transition matrix

Status in T=1 \ Status in T=2	A	B
A	Stayers	Movers (Upward mobility)
B	Movers (Downward mobility)	Stayers

Source: Author

By establishing this type of matrix for each of our classifications of interest, we can compare the mobility patterns of different socio-demographic groups.

Earnings mobility is detected by the variable PC (Percentile Change) which can take the values [-100; 100]. It measures the number of percentiles of mobility that a worker experienced across the distribution of hourly wages between the 2005 and 2011-12.

4.2. Estimation of labor market mobility

Income mobility describes two types of phenomena: an income change during one's lifetime or an income change between generations (e.g. parents and children). This dichotomy is addressed as intragenerational and intergenerational income mobility in the literature. With an intragenerational perspective, this study aims at analyzing the determinants of positional change. Instead of focusing on whether there has been a change in the concentration of individuals in each slot between two periods, our focus lies on the determinants of rank-change across the distribution of earnings in 2005 and 2011-12. This relative approach implies that if everyone benefits from the same share of income increase between the two waves, there is no relative mobility. Conversely, if only one worker experiences a change of income, at least one other person will change places in the distribution.

In contrast with the literature, we shift the focus from incomes to hourly earnings. The perspective adopted in this study is to observe changes in the way a person is remunerated for the same effort, which is better represented by hourly earnings than by monthly earnings or household income. Nevertheless, our focus erases the economic mobility one might benefit from gaining access to more hours of work[51] or to transfers, credit etc., in which case analyzing

[51] Our method does not capture the aspects of mobility consisting in escaping visible under-employment.

monthly or yearly earnings would be more relevant. The form of mobility we analyze should be kept in mind when interpreting the results.

Methods to analyze intragenerational income mobility usually measure transitions in and out of quantiles (often quintiles) of the income distribution (see for instance Ranganathan, Tripathi, and Pandey (2016) in the case of India). The method used by Azam (2016), based on Bhattacharya and Mazumder's (2011) initial work provides fine results concerning the movement dynamics. They analyze transition probabilities as well as upward rank mobility. In a two-period context, upward rank mobility is the probability that a household exceeds a given percentile of the income distribution by an amount τ, conditional on the household's initial position in the distribution. The percentile of reference and τ must be chosen beforehand. Azam (2016) chooses the median as the percentile of reference and 0 and 20 for τ. He also estimates the probability that a household improves its rank by looking at transitions across the median using a linear probability model. We choose to implement a different strategy since an arbitrary choice may not allow analyzing the dynamics of mobility for the whole distribution.

This study takes advantage of the panel dimension of the IHDS dataset to explore the determinants and mechanisms of labor market mobility. In this setup, the different measures of labor market mobility (occupational or earnings) are the dependent variables. Given that we have two types of dependent variables (binary/categorical on the one hand and continuous on the other hand), we will need two types of estimation methods. This section presents the baseline models as well as the estimation strategies that we will use.

4.2.1. Baseline models

Case 1. In the case of occupational mobility, the dependent variable is either mobility as a binary variable (mover *versus* stayer) or a categorical variable with three possible outcomes (upward mobile mover; downward mobile mover; stayer)

$$\Pr(Mobility = 1) = \gamma\left(\beta_j X_i\right) \quad \text{[Eq. 2.1]}$$

With:

- $\Pr(Mobility = 1)$: the probability that the mobility variable takes the value 1 (mobile)
- F: cumulative logistic distribution
- X_i: vector of explanatory variables
- β_j: the vector of coefficients to be estimated.

The alternative specification when *Mobility_multi* is a multinomial variable uses a multinomial logistic estimation:

$$\Pr(Mobility\_multi = 1|X_i) = \frac{\exp(\beta_j X_i)}{1+\sum_{k=1}^{m}(\beta_j X_i)} \qquad \text{[Eq. 2.2]}$$

$$\Pr(Mobility\_multi = 0) = \frac{1}{1+\sum_{k=1}^{m}(\beta_j X_i)} \qquad \text{[Eq. 2.3]}$$

In equations 2.1 and 2.2, X_i contains the group-defining variables (gender, religion and caste), potential predictors of labor market mobility (educational attainment, and indicator of innate ability and the initial earnings *intialY*) and control variables.

Case 2. Our second type of initial equation is an OLS estimation where the dependent variable is the percentile rank change between two given points in time.

To estimate the determinants of earnings mobility, the following equation has to be estimated:

$$PC = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 initialY + \varepsilon \quad \text{[Eq. 2.4]}$$

Where PC is percentile change which can take the values [-100;100], X_1 is a vector of personal characteristics, X_2 is a vector of control variables and *initialY* is the level of earnings in 2005.

4.2.2. Estimation method

Several challenges face the robustness of these initial estimations: (i) the potential sample selection bias and selective attrition bias, (ii) the possible measurement errors in the hourly earnings variable and (iii) simultaneity and unobserved heterogeneity bias.

4.2.2.1. Correcting the potential sample selection bias and selective attrition bias

Two forms of selection bias are likely to affect the results: the sample selection which is due to the fact that the workers in our sample do not represent a random subsample of the all working-age individuals, and the potential non-randomness of panel attrition. In both cases, the OLS estimators will be inconsistent if unobservable factors that affect the inclusion of individuals in the sample are correlated to the outcome of interest.

This type of selection bias can be corrected by using the two-step method developed by (Heckman 1979). It consists in augmenting the equation of interest with the Inverse Mill's Ratio (IMR) computed from the following probit selection equation (Eq. 2.5).

$$\Pr(A_i = 1) = \gamma(x_2 \gamma_2) \qquad \text{[Eq. 2.5]}$$

A_i is a dummy variable which takes the value 0 or 1. x_2 is the vector containing explanatory variables for labor market participation and γ is the normal density function. From this equation, it is possible to compute the inverse Mill's ratio, $\lambda(x_2\widehat{\gamma_2})$ which is added as a regressor in the equations of interest[52]. Note that this method requires specific predictors of A_i, which are not correlated to the outcome of interest (labor market mobility), to be included in equation 2.5. These variables are called exclusion restriction variables.

In order to correct the sample selection bias due to the restriction of the analysis to individuals who work, we use the number of male and female children below five years old and the number of elderly individuals in the household as restriction exclusion variables.

Furthermore, as described in Section 3.2., the potential non-randomness of panel attrition can bias the estimation results. The exclusion restriction variable is supposed to predict whether a person remains in the sample in the second survey round, without predicting mobility. First, following Sarkar, Sahoo, and Klasen (2017) who study female entry and exit in the labor market using IHDS data, we use the person identifier number assigned during the survey as a determinant of whether a person stays in the household between the two rounds. The identification relies on the assumption that individuals who answer the questions are recorded first. Indeed, they are the most likely to be attached to the household and therefore to remain in the sample in the following survey wave. We also add a binary variable indicating pension benefits and the number of days a person was ill because of a major morbidity issue, which influences their probability of remaining in the second wave of the survey. Migration is also probably an important determinant of the probability of remaining in the sample but it is also a potential predictor of occupational and/or earnings mobility. Moreover, because of missing data, we cannot include it in our specifications.

4.2.2.2. Accounting for measurement error in earnings: a simulation exercise

Since the dependent variable PC is computed from reported earnings in 2005 and 2011-12, it is likely to be affected by measurement error, which is all the more probable given the focus on hourly earnings computed by combining information on salary and working hours. Studies in development economics usually assume that the measurement error linked to misreported earnings is a *classical measurement error* issue. This implies that when doing a regression in

[52] A discussion on selection into the labor market and an alternative method to correct this bias is presented in Chapter 3.

which the earnings variable is an exogenous variable, the measurement error will affect the standard errors but the coefficients will be unbiased. In this case, robust standard errors can be derived from a bootstrapping method[53] which consists in replicating the equation in random samples drawn from the initial sample, and calculating standard errors based on all of the replications.

To ensure that our estimations are robust to this potential bias, the standard errors are computed using a simulation exercise in which we bootstrap not only *the regressions*, but also *the generation of the distributions of earnings* and *the derivation of the Percentile Change variable*. This process ensures more robust distributions of hourly earnings. The rationale of our methodology is that by randomly dropping individuals from the sample throughout multiple replications, it is likely that the misreported earnings are dropped. Therefore, the two distribution functions (one for 2005 and one for 2011-12) and calculate the percentile change in each replication. Consequently, the mobility we detect is less likely to be spurious and the problem of misreporting earnings is minimized.

Each replication follows step 1 to 3:

1. Generate a distribution of hourly earnings for 2005 and 2011-12
2. Calculate the percentile change (*PC*) between both waves for each individual
3. Estimate the determinants of mobility[54]

Nichols (2010) uses a bootstrapping method to estimate mobility risk. However, adding the derivation of the percentile change (PC) in the bootstrapping process is a methodological contribution of this study. The main caveat of the methodology described above is that it assumes that the errors in the declaration of earnings are distributed randomly in the population. Indeed, each draw of the simulation randomly excludes individuals. It would be possible to condition the draws not to be random, but this would require knowing which variables are correlated to the measurement error and in which way. A possible reason for the bias would be income shocks that influence an individual's ability to recall their exact earnings. Nonetheless, Akee (2011) finds that in Micronesia, this type of error tends to disappear in the first or second

[53] In microeconometrics, bootstrapping methods are usually used to compute robust standard errors when two or more equations are successively used to estimate a phenomenon, and one equation uses values (such as a coefficient, or an error term) computed from the other equations. For instance, a common method that requires bootstrapped standard errors is the computation of the Inverse Mill's Ratio in a two-step Heckman selection correction method.

[54] This step includes the Heckman selection correction steps to correct for non-random labor market participation and selective attrition.

year following the shock. Other possible reasons for non-classical measurement errors would be that individuals are not as likely to correctly report their earnings based on their education level and their ability in basic algebra. The lack of studies on this type of bias in the case of India makes it impossible for us to detect and correct it. More generally, Akee (2011) suggests that there is no solution to the issue of misreported errors apart from better measurement or finding instruments for income when they are used as an independent variable. We will, therefore, assume that measurement errors in the dependent variable are randomly distributed.

4.2.2.3.Dealing with the endogeneity of *InitialY*

An endogeneity bias stems from the initial level of hourly income (*initialY*) variable, which is a determinant of both occupational mobility and earnings mobility. Its inclusion in the estimations as an independent variable is likely to yield biased estimates. The main caveat of our methodology, if we decided to exclude the *InitialY* variable would be to consider movements across the distribution to be the same, regardless of the initial income. It is therefore essential to include this variable in the specifications.

In the first set of specifications concerning occupational mobility (**case 1**), this variable is prone to an unobserved heterogeneity bias. For instance, we can suppose that individuals who are more likely to experience occupational mobility are the ones who earned more to begin with, for reasons related to ability or motivation. It is possible to correct this bias with an Instrumental-Variable (IV) method. Using IV to estimate multinomial logistic models requires implementing a control function approach (Petrin and Train 2010; Wooldridge 2015) by estimating the two following equations.

$$initialY = \delta_0 + \delta_1 instrument + \delta_2 X_2 + \mu \quad \text{[Eq. 2.6]}$$

$$\Pr(Mobility\_multi = 1 | X_i, \widehat{\mu_i}) = f(X_i, \widehat{\mu_i}) \quad \text{[Eq.2.7]}$$

In a first step we estimate the endogenous covariate *initialY* with an Ordinary Least Squares estimation, next we input the vector of reduced-form errors $\widehat{\mu_i}$ into the multinomial logistic estimation[55].

In the second type of specification concerning earnings mobility (**case 2**), an important determinant of percentile change is probably the level of earnings at the beginning of the

[55] These two-steps are bootstrapped with 500 replications in order to obtain consistent estimated. Note that this bootstrap process is not the same as the 3-step method we use to control for possible misreporting of earnings.

observation period. In intergenerational studies, the coefficient of parents' income variable in the estimation of a person's income indicates the extent to which earnings are "*sticky*" across generations. If this coefficient is high, then a parent's place in the distribution is a good indicator of a child's place. When this coefficient is high the stickiness of earnings is therefore considered as substantial (Corak 2013). We can easily assume that this trend is also valid for individual trajectories in the medium-run. A person who initially has higher earnings might not have the same chances of mobility than a person with lower earnings. More specifically the endogeneity bias can be summed up in the two following ways:

- a simultaneity bias since *initialY* is used to calculate the variable *Percentile Change*.
- an unobserved heterogeneity bias since the error term might be correlated to *PC*. The variation of *initialY* might be correlated to the variation in PC because of other factors which cannot be controlled for such as motivation or soft skills.

Therefore, we will implement a Two-Stage Least Squares (2SLS) method using instrumentation as an attempt to correct this endogeneity bias. The following equations are estimated consecutively.

First-stage equation:

$$initialY = \delta_0 + \delta_1 instrument + \delta_2 X_2 + \mu \quad \text{[Eq. 2.8]}$$

Second-stage equation:

$$PC = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \widehat{initial}Y + \varepsilon \quad \text{[Eq. 2.9]}$$

In both cases described above (**case 1** and **case 2**), finding relevant instruments implies finding one or several variables that are correlated to *initialY* without being correlated (or with the less possible correlation) to labor market mobility. We use the two following variables as instruments. The first one is a household asset score ranging from 0 to 30 in 2005[56]. Although there is a possibility that this instrument is correlated to labor market mobility, we argue that in the medium-run, household assets can only minimally influence one's probability of changing their place in the earnings distribution among salaried workers. This is more likely among self-employed individuals and household businesses, which are not included in this study. Among

[56] Assets include: sewing machine, mixer/grinder, motor vehicle, TV, air cooler, air conditioner, electric fan, chair/table, cot, telephone, cell phone, refrigerator, pressure cooker, any vehicle, car, clock/watch, washing machine, computer, credit card, 2 clothes, footwear, piped indoor water, separate kitchen, flush toilet, electricity, solid (mentioned as *Pucca* in the IHDS questionnaire) wall, roof and floor.

these households, there is an unclear boundary between household assets and the investments in their productive activity. Second, following the recommendation of (Wooldridge 2010) and implementations in several studies (see for instance Dumas (2012)) which use community-level variables as instruments, we compute the community education level of the household head (in 2005) at the PSU level which is an indication of labor supply characteristics.

To sum up our methodology, we propose to bootstrap the following process to estimate earnings mobility:

1. Generate a distribution of hourly earnings for 2005 and 2011-12
2. Calculate the percentile change (*PC*) between both waves for each individual
3. Estimate the determinants of mobility
    a. Calculate the selection terms from the Heckman equations
    b. Estimate earnings mobility with a 2SLS method

4.2.3. Estimating earnings mobility using an alternative dependent variable

Since the endogeneity of equation 2.4 comes from the value of initial earnings, one way of preventing this endogeneity is to transfer the *initialY* variable to the left-hand side of equation 4. In order to do this, we reweight PC by the log of earnings. This provides us with a variable that is a weighted percentile change (*WPC*). For the same level of rank change, a higher income will yield a smaller *WPC*, and a lower income will yield a larger WPC. Equation 2.10 is estimated with OLS following the bootstrapping process described in Section 4.2.2.2.

$$\mathrm{W}PC = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad \text{[Eq. 2.10]}$$

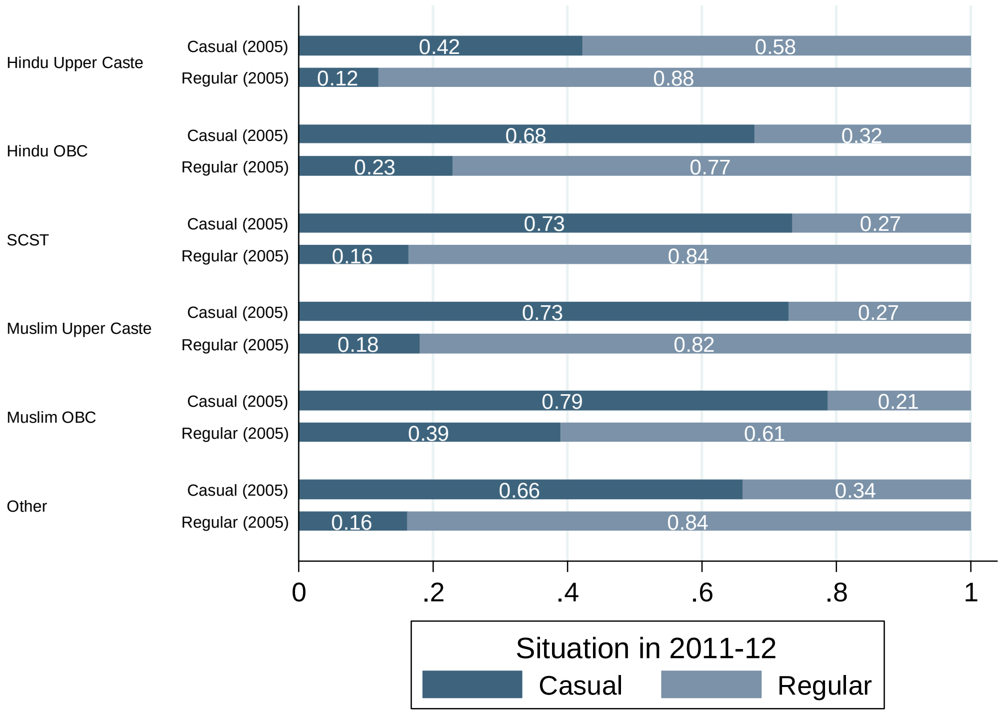
5. Patterns of labor market mobility

This section describes the patterns of labor market mobility in regarding occupational transition (5.1) and relative earnings mobility (5.2).

5.1. Mobility matrices

The mobility matrices presented in section 4.1. are presented in graphical form (Figures 2.2 to 2.7). They show the extent of mobility for each type of occupational status by socio-religious groups or by gender. The distribution of gender, religion and caste groups across the occupational status are presented in Appendix 2.1. The main trends that can be observed from the distribution of individuals in different occupational groups in 2005 are that the share of men and Hindu Upper Castes is systematically higher in regular and higher skilled occupations. In the public administration and services, Hindu Upper Castes have the highest share of workers (37.7%). Concerning the manufacturing sector, OBCs have the highest share of workers (37.0%), and in the urban agricultural sector and in construction SCSTs have the highest shares of workers.

Figure 2.2. Casual-Regular occupational transition by religion and caste between 2005 and 2011-12



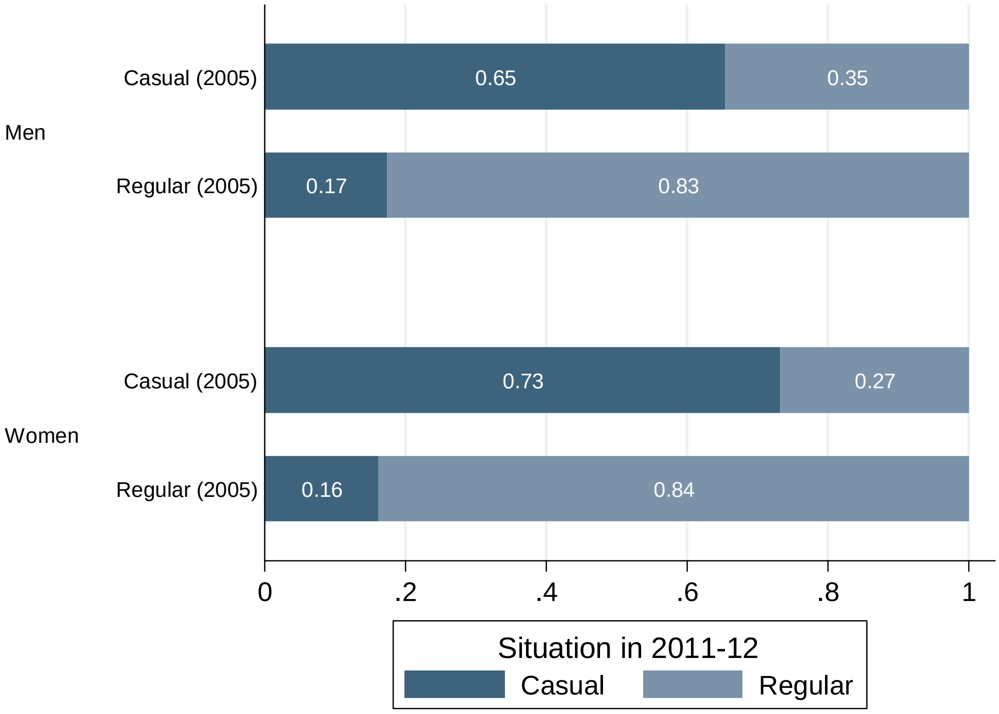
Source: Author's calculations from the IHDS dataset

Figure 2.2 presents transitions from casual to regular occupations by religion and caste groups between 2005 and 2011-12. Each row represents an occupational status in 2005. The colors and statistics in the row indicate the occupational status in 2011-12.

Assuming that casual-to-permanent occupational mobility is an upward form of mobility, a considerable level of upward occupational mobility is visible in urban India from 2005 to 2011-12. Indeed 33.37% of casual workers transitioned to regular occupations. Although these occupations are not necessarily formal, they are generally associated with better working conditions than casual labor. These forms of employment are also less likely to be categorized as time-related underemployment, which makes them potentially more productive. The Hindu

Upper Caste group has a higher upward mobility rate than the other groups. Indeed, 58% of initially casual workers from this group transitioned to regular occupations in 2011-2012, whereas this share does not exceed 32% for the other groups. The Muslim OBC community is the most concerned by downward mobility with 39% of regular workers in 2005 who transitioned to casual jobs in 2011-12.

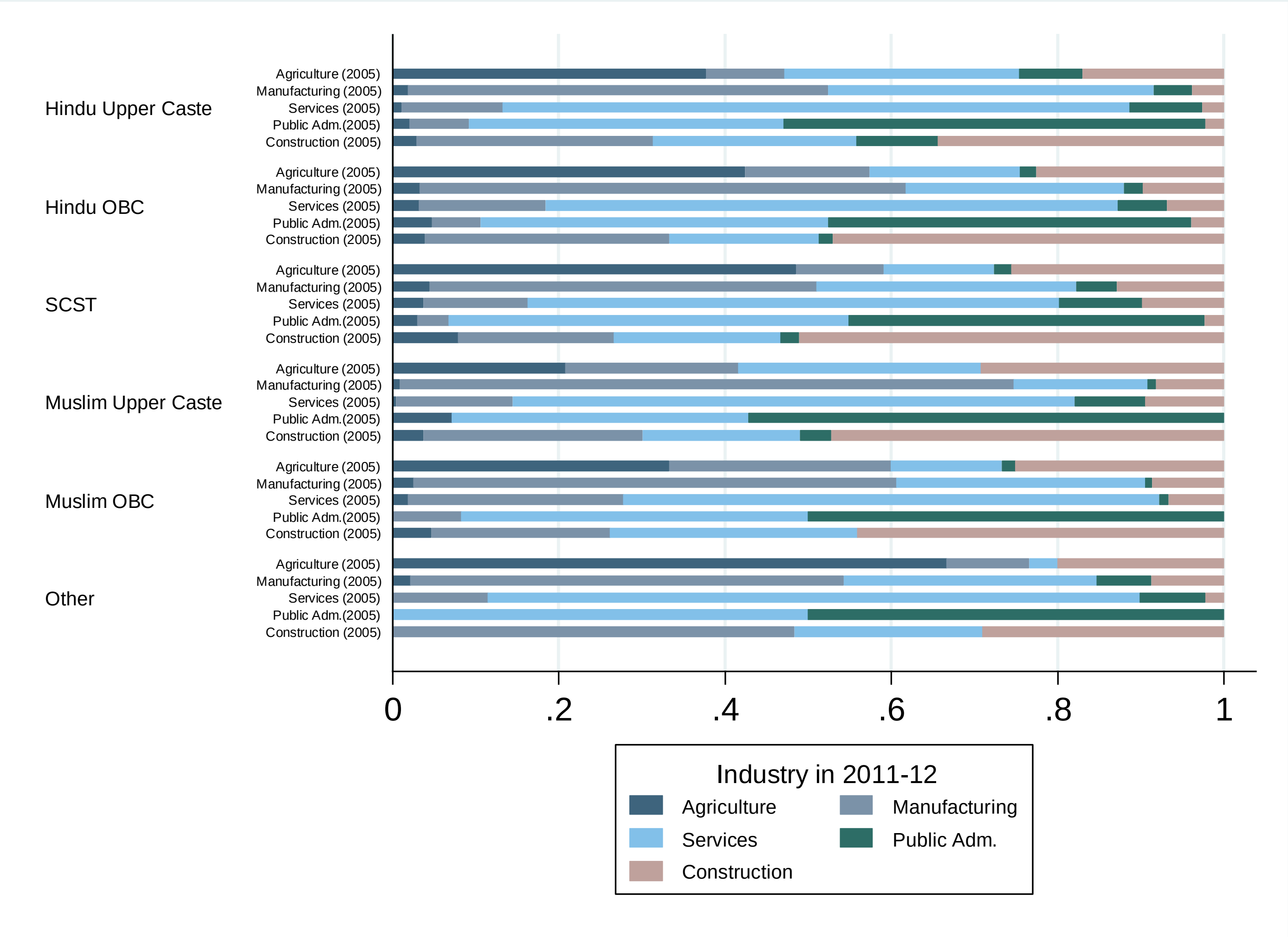
Figure 2.3. Casual-Regular occupational transition by gender between 2005 and 2011-12



Source: Author's calculations from IHDS data

Women benefit from less upward mobility than men (27% and 35% respectively), but the levels of downward mobility are very close across both groups. It is possible that women experience lower mobility because of a lack of opportunities for them in regular occupations.

Figure 2.4. Transition across industries by religion and caste between 2005 and 2011-12



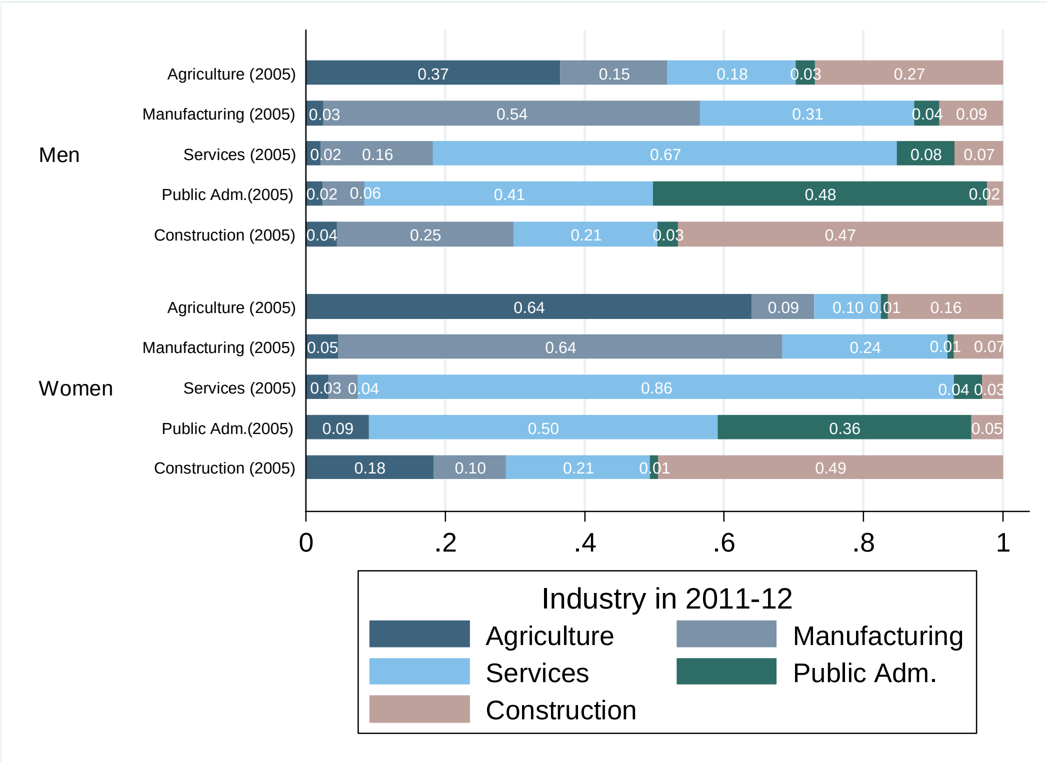
Source: Author's calculations from IHDS data

The inter-industrial mobility patterns presented in Figure 2.4 are very interesting in terms of caste and religion as it provides us with a first insight on path-dependency in the short-term.[57] Workers are classified in five industrial sectors: Agriculture, Manufacturing, Services, Public Administration and Construction. Among Hindu Upper Caste workers, there is a majority of service workers (65.8% in 2005 and 64.1% in 2011-12). The transition graph also shows important transitions into the services sector for Hindu Upper Caste workers who were in other sectors in 2005. 28% of agricultural workers, 39% of manufacturing workers, 38% of public administration workers and 25% of construction workers from this socio-religious group transitioned into the services sector. 75% of the Hindu Upper Caste service workers remained in the same industry between the two dates ("*stayers*"). Stayers in the services sector also represent important shares of workers in the other groups (69% for the Hindu OBC group, 64% for SCSTs, 68% for Muslim Upper Castes and 64% of Muslim OBCs). The group whose share of stayers in the manufacturing sector is the highest is Muslim Upper Caste (74%), followed by Hindu OBC and Muslim OBC (58% for both groups). The graph also shows considerable mobility in and out of the construction sector.

Transitions are visible for all of the groups, indicating that caste-based occupational segregation is not a rigid reality. These patterns, especially generalized entry into the services industry, seem to indicate that the development of the services sector is contributing to occupational mobility by caste. What is not visible, however, is how the groups are distributed inside the services sector. Indeed, it is plausible that the group hierarchy is maintained in this modern sector, lower castes occupying more precarious forms of employment. This would attest to the flexible nature of the caste system which can create new forms of segregation.

[57] The detailed figure, which indicates the shares of the transitions is presented in Appendix 2.2.

Figure 2.5. Transition across industries by gender between 2005 and 2011-12

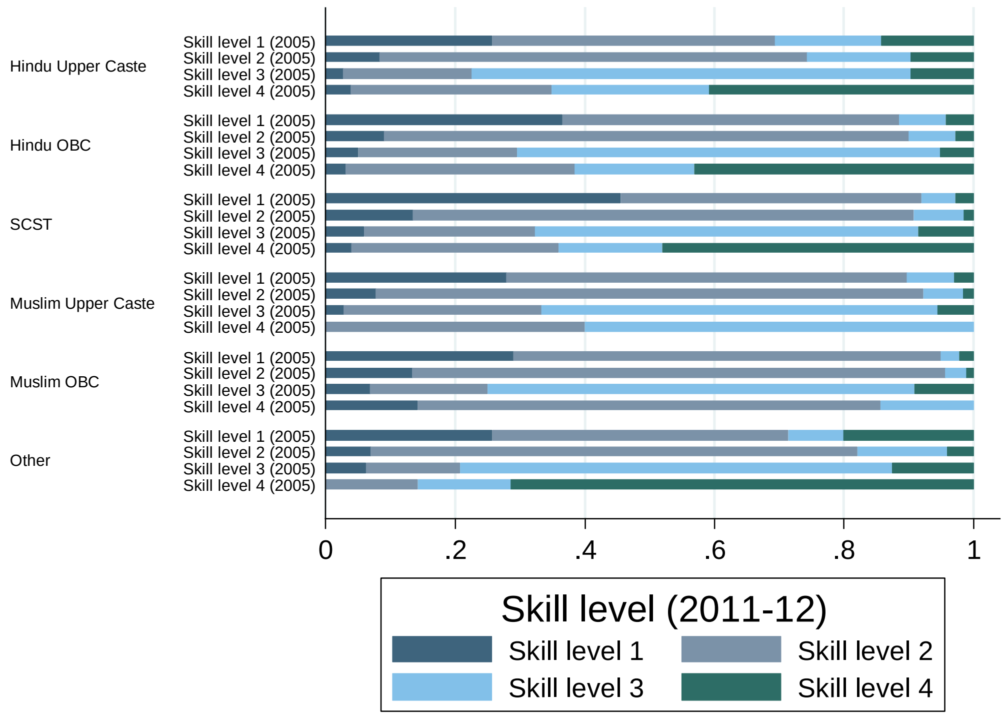


Source: Author's calculations from IHDS data

Figure 2.5 shows inter-industrial transitions by gender groups and suggests that women are more path-dependent than men since they systematically show higher levels of non-mobility across industries. The most striking difference concerns the Agricultural sector in which there are 37% of male stayers and 64% of female stayers.[58]

[58] Note that this study is restricted to urban areas, in which agriculture takes the form or urban or peri-urban farming.

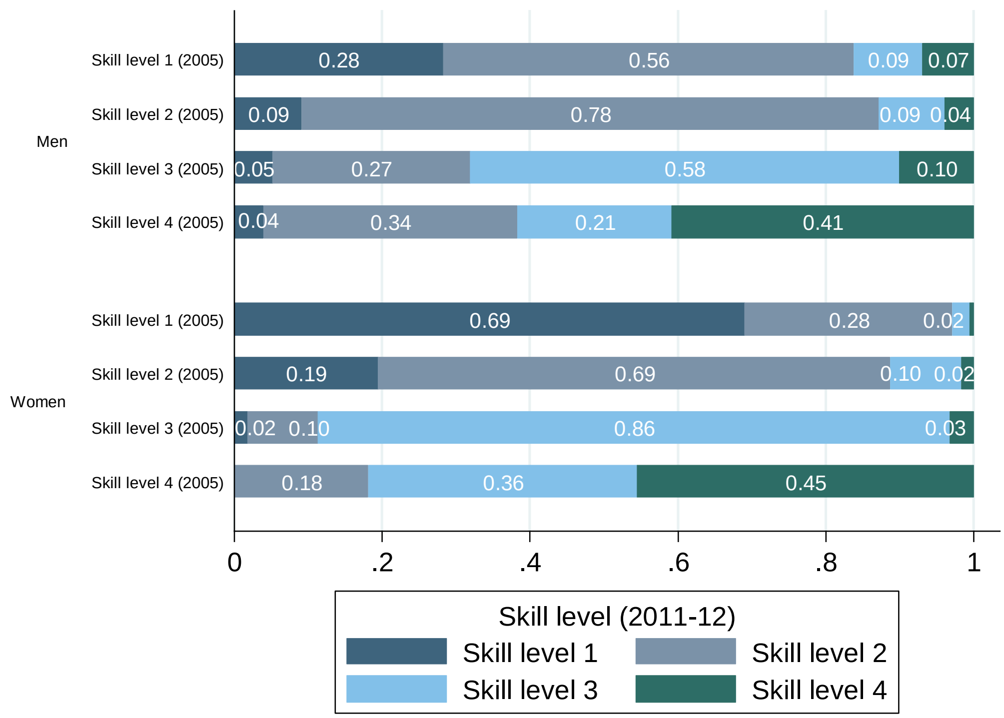
Figure 2.6. Transition across skill levels in occupations by religion and caste between 2005 and 2011-12



Source: Author's calculations from IHDS data

Figure 2.6 shows the patterns of upward skill mobility by religion and caste. Upward skill mobility for workers that started with the Skill level 1 is more limited for the SCST group since they have the highest share of stayers. Transitions out of this skill level are higher for workers from the Hindu Upper Caste group. In all cases, workers who exit Skill level 1 occupations predominantly enter Skill level 2 occupations. Overall, workers from Skill level 2 have a higher probability of immobility. The previous discussion about the caste hierarchy can be completed by the findings on skill mobility. If skill mobility patterns are the same for all industries, SCSTs are potentially less likely to gain access to the skilled occupations of the service sector.

Figure 2.7. Transition across skill levels in occupations by gender between 2005 and 2011-12



Source: Author's calculations from IHDS data

Figure 2.7 shows considerable differences in women's probability of staying in Skill 1, 3 and 4. Women have lower levels of upward skill mobility (more stayers in level 1 and level 2). Interestingly, the fact that they have a smaller share of stayers in Skill level 2 is not linked to upward but to downward mobility. Indeed, both groups face important levels of downward mobility from Skill level 4 to 3.

To sum up the results from the mobility matrices, Upper Caste Hindus seem to benefit from higher upward mobility (from casual to regular, out of skill level 1). SCSTs face higher skill level 1 immobility compared to Hindu OBCs and Muslims. Muslim OBCs face higher immobility in casual employment compared to Hindu OBCs, SCSTs and Muslim Upper castes. Industrial mobility is important for all groups, especially into the services sector. Regarding gender, the mobility matrices show that the share of stayers is higher for women than men in casual employment, all of the industries and all skill level, except for Skill level 2 in which case they face considerable downward mobility. These results show important differences in the patterns of occupational mobility. Section 5.2 presents the descriptive statistics for hourly earnings mobility.

5.2.Relative earnings mobility

Table 2.5 shows the descriptive statistics of Percentile Change between 2005 and 2011-12. Appendix 2.3 shows the hourly earnings per year and group.

Table 2.5. Percentile change

Variable	N	Mean	Standard Deviation
	Whole sample		
Percentile Change	7,024	6.036	23.711
Percentile Change<0	2,420	-17.710	16.425
Percentile Change>0	4,379	19.470	16.086
	Female subsample		
Percentile Change	1,040	5.565	23.25
Percentile Change<0	358	-15.989	16.759
Percentile Change>0	637	18.072	17.465
	Male subsample		
Percentile Change	5,984	6.118	23.79
Percentile Change<0	2.062	-18.008	16.352
Percentile Change>0	3,742	19.707	15.830
	Hindu Upper Caste		
Percentile Change	1,670	2.746	23.389
Percentile Change<0	633	-18.376	18.167
Percentile Change>0	956	16.965	15.455
	Hindu OBC		
Percentile Change	2,146	6.882	24.007
Percentile Change<0	733	-17.484	15.340
Percentile Change>0	1,355	20.358	16.752
	SCST		
Percentile Change	1,974	7.278	23.345
Percentile Change<0	637	-17.123	16.243
Percentile Change>0	1,273	19.855	15.945
	Muslim Upper Caste		
Percentile Change	414	6.370	24.315
Percentile Change<0	141	-18.865	17.008
Percentile Change>0	264	20.064	15.600
	Muslim OBC		
Percentile Change	519	7.287	23.259
Percentile Change<0	170	-18.000	15.568
Percentile Change>0	341	20.065	14.783

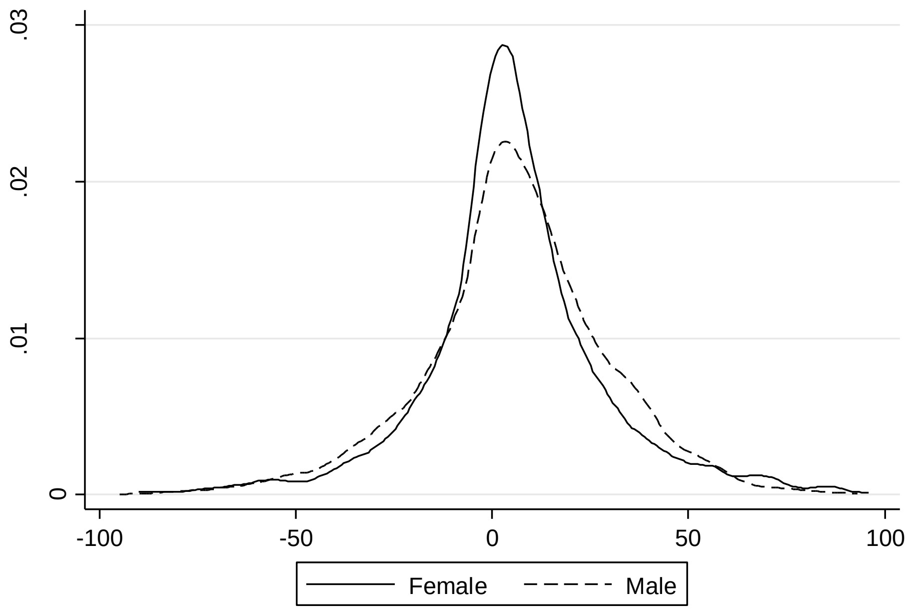
Source: Author’s calculations from the IHDS dataset
Note: Student’s T-tests shows non-significant differences for men and women. Significant differences are found for all caste groups compared to the rest of the population.

A positive rank change in the distribution of hourly earnings (PC>0) can have two meanings, either the person experienced an increase in hourly earnings, or at least one other person experienced a negative change in hourly earnings. On average, an earnings mobility of 6 percentiles is visible between 2005 and 2011-12. 34% of individuals experienced downward mobility (PC<0) with an average of -17.710 percentiles change. The remaining 66% experienced upward mobility (PC>0) with an average of 19.470 percentile change. Percentile change is not significantly different by gender. Regarding socio-religious groups, Hindu Upper Castes have significantly lower levels of mobility compared to the other groups (on average 2.746 percentile change) while SCSTs and Muslim OBCs have seen relatively higher mobility with approximately seven percentiles of upward mobility.

The kernel density plots (Figure 2.8. and 2.9)[59] show that the most frequent rank change is around 0 which means that for most workers, the earnings growth or earnings loss was not significant enough to change their place substantially in the distribution.

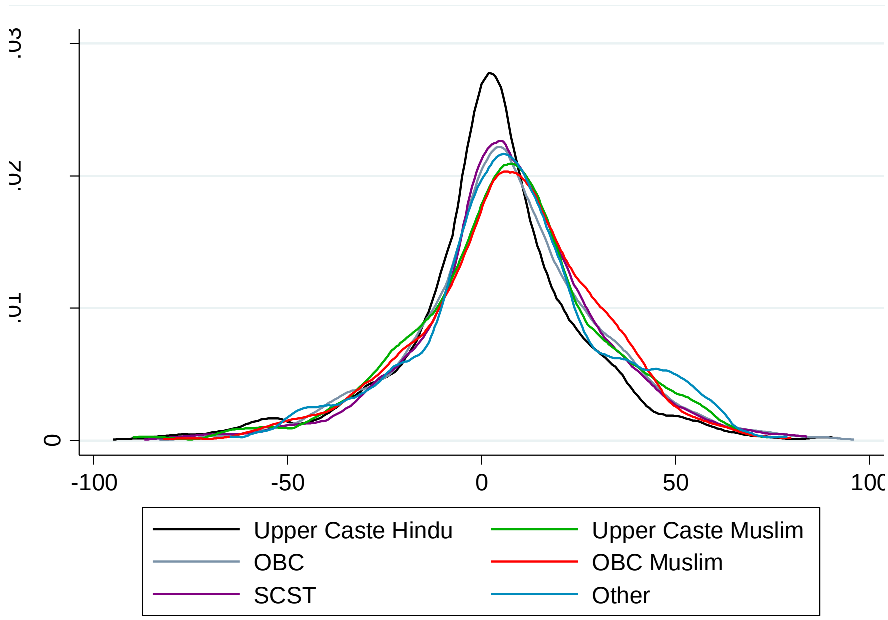
[59]Histograms for each group are presented in Appendix2.4.

Figure 2.8. Kernel density plot of Percentile Change by gender between 2005 and 2011-12



Source: Author's calculations from IHDS data

Figure 2.9. Kernel density plot of Percentile Change by religion/caste between 2005 and 2011-12



Source: Author's calculations from IHDS data

The male distribution in Figure 2.8 has a larger right-side tail, until approximately the 60th percentile change, suggesting that men experienced more upward mobility than women (except for a few women who experienced high mobility). The distribution for men also has a larger left-side tail, which shows that men experienced more downward mobility. Overall, women seem to have experienced less upward or downward mobility.

The sample disaggregated by religion and caste show a more leptokurtic distribution for the Hindu Upper Caste group and more platykurtic distributions for both Muslim groups. The distribution for Hindus Upper Caste stands out the rest of the population. Indeed, the distributions of percentile change for Hindu OBCs, SCSTs and Muslims are all more left-skewed and have larger right tails than for Upper Caste Hindus. These trends suggest that Upper castes have known less upward mobility than the other groups. Given their higher social status, the other groups apprear to be in the process of "*catching up*."

6. The determinants of mobility

This section presents the estimation results of labor market mobility. Table 2.6 recapitulates the dependent variables of the estimations and Appendix 2.5. shows the means and standard deviations of the variables used in the estimations. They are all from the first wave of data and include gender, religion and caste, highest educational attainment, the number of children, whether the person is married, a State control variable[60] and the classification of the worker in the secondary school leaving certificate (SSLC) examination. This certificate is necessary for pursuing schooling and individuals are classified in three groups, from highest to lowest level of distinction Class I (*Sec1*), Class II (*Sec2*) and Class III (*Sec3*). Individuals who did not pass the SSC examination or have a missing value are coded 0. This variable has been used to control for innate ability by Azam, Chin, and Prakash (2013) and Sahoo and Klasen (2018).

[60] Note that alternative specifications including a district-specific price of a goods basket, aiming at capturing local economic constraints, did not show significant differences from the main results presented in this section. It was therefore excluded from the presented estimations to avoid any form of "bad control" variable (Angrist and Pischke 2009).

Table 2.6. Variable description

Dependent Variable	Variable Description
Professional Mobility Variable	
Casual-Permanent	0: No Mobility
	1: From regular (2005) to casual (2011-12)
	2: From casual (2005) to regular (2011-12)
Mobility Industry	0: No mobility
	1: Mobility
Mobility Occupation	0: No mobility
	1: Downward skill mobility
	2: Upward skill mobility
Hourly Earnings Percentile Change	
Percentile Change	Percentile change in the distribution of hourly earnings

Source: Author's calculations from IHDS data

6.1. Determinants of occupational mobility

Table 2.7 presents the results of the multinomial logistic and logistic estimations of occupational mobility using a control function approach to correct the endogeneity linked to initial earnings, with selection correction terms added to control for non-random labor market participation (*Selection_correction*) and non-random panel data attrition (*Selection_attrition*). The probit equations used to calculate these selection terms are presented in Appendix 2.6. The results without the control function are presented in Appendix 2.7.

Table 2.7. Professional mobility estimations[61]

	Estimation 1		Estimation 2	Estimation 3	
	Casual-Regular mobility (ref. group: No mobility)		Industrial mobility (ref. group: No mobility)	Skill levels in occupations mobility (ref. group: No mobility)	
VARIABLES	Downward	Upward	Mobile	Downward	Upward
Female	-0.147	0.077	0.039	-0.054	**0.513*** **
	(0.177)	(0.100)	(0.057)	(0.110)	**(0.066)**
Hindu OBC	0.021	-0.025	**0.184*** **	-0.100	0.012
	(0.176)	(0.101)	**(0.069)**	(0.120)	(0.071)
SCST	-0.134	**-0.490*** **	0.109	0.127	**-0.156** **
	(0.214)	**(0.123)**	(0.077)	(0.133)	**(0.079)**
Muslim Upper Caste	-0.084	**-0.412** **	**0.212** **	-0.237	**0.202** **
	(0.307)	**(0.160)**	**(0.105)**	(0.195)	**(0.099)**
Muslim OBC	0.140	**-0.485*** **	**0.461*** **	0.227	**0.398*** **
	(0.269)	**(0.152)**	**(0.101)**	(0.177)	**(0.098)**
Other	-0.103	**-0.394* **	0.114	-0.386	-0.016
	(0.377)	**(0.210)**	(0.122)	(0.242)	(0.131)
Educ_primary (2005)	-0.047	**0.324** **	0.148	0.230	0.046
	(0.246)	**(0.139)**	(0.092)	(0.153)	(0.090)
Educ_middle (2005)	0.162	**0.588*** **	0.053	0.077	**0.199*** **
	(0.152)	**(0.097)**	(0.061)	(0.115)	**(0.066)**
Educ_secondary (2005)	0.022	**0.942*** **	0.104	0.259	**0.301** **
	(0.378)	**(0.186)**	(0.132)	(0.222)	**(0.138)**
Educ_higher (2005)	0.072	**0.565*** **	0.008	**0.498** **	**0.485*** **
	(0.416)	**(0.209)**	(0.142)	**(0.239)**	**(0.162)**
Sec1 (2005)	-0.283	**-0.633*** **	-0.131	0.215	0.266
	(0.432)	**(0.223)**	(0.148)	(0.250)	(0.163)
Sec2 (2005)	0.023	-0.265	-0.008	0.162	0.103
	(0.376)	(0.187)	(0.132)	(0.223)	(0.141)
Sec3 (2005)	0.248	-0.062	-0.074	0.306	0.023
	(0.437)	(0.216)	(0.157)	(0.236)	(0.159)
Age (2005)	**0.195** **	-0.022	**-0.091*** **	**0.084* **	-0.030
	(0.087)	(0.048)	**(0.030)**	**(0.050)**	(0.029)
Age squared (2005)	**-0.002** **	-0.000	**0.001*** **	-0.001	**0.001** **
	(0.001)	(0.001)	**(0.000)**	(0.001)	**(0.000)**
Number of children (2005)	-0.073	0.015	**-0.036** **	**-0.084** **	**-0.045** **
	(0.049)	(0.026)	**(0.016)**	**(0.034)**	**(0.018)**
Married (2005)	0.029	**-0.243** **	-0.071	-0.127	-0.004
	(0.166)	**(0.095)**	(0.066)	(0.117)	(0.063)

Table 2.7 continued on next page

[61] Alternative estimations adding a dummy variable for individuals who have multiple jobs show very similar results.

Table 2.7 (continued)					
Selection_correction	1.186*	-0.131	0.022	**0.885****	**0.650*****
	(0.640)	(0.355)	(0.218)	**(0.401)**	**(0.231)**
Attrition_correction	-0.429	**0.462***	**-0.505*****	0.014	**-0.600*****
	(0.496)	**(0.243)**	**(0.167)**	(0.301)	**(0.185)**
InitialY	**0.028****	**0.000***	0.001	0.011	**0.040*****
	(0.014)	**(0.008)**	(0.005)	(0.009)	**(0.005)**
$\widehat{\mu_i}$	-0.006	**-0.061*****	**-0.004*****	-0.000	**-0.005*****
	(0.005)	**(0.004)**	**(0.002)**	(0.003)	**(0.002)**
State control variables	Yes		Yes	Yes	
Constant	**-8.079****	-0.153	**2.096****	**-4.033*****	-0.831
	(3.206)	(1.331)	**(0.859)**	**(1.448)**	(0.849)
Observations	6807		9735	9735	

Source: Author's calculations from the IHDS dataset
Notes: Bootstrapped standard errors in parentheses (500 replications)[62], *** p<0.01, ** p<0.05, * p<0.1

The description of the mobility matrices in Section 5 shows that there is a larger path-dependency (no mobility) for women than for men between casual and regular occupations. However, after controlling differentials linked to education, productivity levels and other characteristics, gender is not a determinant of casual-regular mobility. This result suggests that transitions into regular employment, which is very small for women, is mostly due to productivity-related factors such as education or innate ability. The results also show that SCSTs, Muslim Upper Castes and Muslim OBCs have a significantly lower chance of experiencing casual-to-regular mobility compared to Upper Caste Hindus (by 38.7%, 34% and 38.5% respectively).[63] Compared to "*no mobility*," upward mobility is mostly driven by productivity-related factors. Upward mobility is 38.3% more likely for individuals with a primary education level than for individuals with no formal education. The likelihood increases up to 80% for individuals with a middle school education level and to 156% for individuals with a secondary level. The only variables that significantly affect regular-

[62] Although the dependent variables in these estimations are not prone to the bias due to earnings being misreported, we still provide the bootstrapped standard errors because the estimation is done in several steps to add the correction terms computed from the Heckman models.

[63]The Relative Risk Ratios (RRR) are calculated by exponentiating the coefficient. For instance, the RRR for SCSTs is calculated in the following way: $e^{0.490} = 0.613$. If the RRR is higher than 1, it means that the SCST group is more likely to experience upward mobility than the reference group (Hindu Upper Caste). If the RRR is smaller than 1 it means that the SCTS is less likely to experience upward mobility than the reference group. Here, the RRR=0.613 which means that the SCST group is 38.7% (=1 - 0.613) less likely to experience upward mobility than the reference group.

to-casual mobility are age and age-squared. As a worker's age increases, the likelihood of experiencing this form of downward mobility increases as well.

Two contrasted assumptions can be made to understand industrial mobility. Either it reflects the ability for a worker to seize an opportunity in another industry, or it reflects a more insecure labor market where workers are vulnerable and experience an industry change as a consequence of unemployment shocks. The fact that education levels or the innate ability proxies (*Sec1, Sec2 and Sec3*) are not correlated to industrial mobility either support the second assumption, or implies that mobility across industrial sectors to seize a new opportunity is driven by other factors such as having the right type of social network. In comparison to being from a Hindu Upper Caste, a Hindu OBC worker is significantly more likely to experience industrial mobility by about 20.2%. This likelihood is also significant and positive for Muslim Upper Castes (23.6%) and Muslim OBCs (58.6%). Furthermore, the fact that Upper Caste Hindus have a lower chance of changing industries is probably linked to forms of upper caste monopoly. This group tends to occupy niches of the economy with important roles of social networks that encourage within-caste turnover and limit the access to these occupations for other groups (Deshpande 2003).

The estimations in Table 2.7 show that mobility between skill levels is less likely for SCSTs and more likely for Muslim OBCs. Furthermore, the relative chances for a woman to experience upward mobility in terms of skills in an occupation are 67% higher than for men. Note that the mobility matrices showed that women were systematically less mobile than men in terms of skills. This result, therefore, implies that women's upward mobility is more likely to be higher compared to men's, when other factors related education and ability are controlled for. The results in Appendix 2.7 show that without the correction for the endogeneity of initial earnings, the effect is negative. When all other factors are held constant (including initial earnings), women are more likely to experience upward skill mobility. It is possible that the lack of equal education or the fact that they have lower levels of earnings to begin with impedes on women's skill mobility. This is the only occurrence for which gender is a significant determinant of job-related mobility. Overall, education level is an important determinant of occupational mobility in terms of skills.

6.2. Determinants of percentile rank change

6.2.1. Main results

Table 2.8 presents the results of different estimations of percentile rank change (PC). Estimations 4 is the OLS estimation; estimation 5 is the bootstrapped OLS estimation in which the distributions of hourly earnings for each year are generated in each bootstrap replication. Estimation 6 shows the Two-Stage Least Square estimation which corrects the endogeneity linked to initial earnings[64]. Estimations 7, 8 and 9 combine the 2SLS method and the bootstrapping method, ensuring that the coefficients and standard errors are also robust to measurement error. All estimations contain the selection correction terms linked to non-random labor market participation and selective attrition.

Table 2.8. Percentile rank change estimations

	Estimation 4	**Estimation 5**	**Estimation 6**	**Estimation 7**	**Estimation 8**	**Estimation 9**
	OLS	**Bootstrapped OLS**	**2SLS**	**Bootstrapped 2SLS**		
	Percentile change	**Percentile change**	**Percentile change**	**Percentile change**	**Percentile change**	**Percentile change**
					(PC<0)	**(PC>0)**
VARIABLES						
initialY	**-3.534*** **	**-0.514*** **	**-0.159** **	**-0.241** **	0.110	-0.001
	(0.812)	**(0.034)**	**(0.0679)**	**(0.098)**	(0.131)	(0.090)
Female	**-1.862** **	**-2.265** **	**-2.307*** **	-1.100	-0.083	**2.493***
	(0.850)	**(1.056)**	**(0.856)**	(1.231)	(1.272)	**(1.346)**
Hindu OBC	0.0872	-1.886	-1.289	-1.671	1.331	-0.516
	(0.992)	(1.231)	(0.864)	(1.269)	(1.229)	(1.250)
SCST	-0.786	-0.055	0.108	0.232	-0.786	-0.185
	(1.361)	(1.422)	(1.004)	(1.444)	(1.418)	(1.509)
Muslim Upper Caste	-1.301	-1.064	0.134	1.444	0.215	1.145
	(1.218)	(1.851)	(1.407)	(1.911)	(1.856)	(2.156)
Muslim OBC	1.560	-0.742	-0.625	-1.002	-0.380	0.254
	(1.597)	(1.763)	(1.248)	(1.751)	(1.542)	(1.936)
Other	-0.485	2.830	1.320	2.572	-0.653	-1.329
	(0.396)	(2.252)	(1.588)	(2.264)	(2.164)	(2.394)

Table 2.8 continued on the next page

[64] Note that the Durbin-Wu-Hausman test for endogeneity rejects the exogeneity of the initialY at the 1% significance level.

Table 2.8 (continued)						
Educ_primary (2005)	0.005	-0.639	**-1.095*****	**-1.161****	**-1.362****	-0.165
	(0.005)	(0.546)	**(0.421)**	**(0.591)**	**(0.559)**	(0.615)
Educ_middle (2005)	0.589	0.007	**0.0113****	**0.013***	**0.016****	0.001
	(1.125)	(0.007)	**(0.00497)**	**(0.007)**	**(0.007)**	(0.007)
Educ_secondary (2005)	0.910	0.021	-0.297	-1.250	-0.522	-2.524
	(0.790)	(1.568)	(1.178)	(1.677)	(1.601)	(1.806)
Educ_higher (2005)	**4.721*****	-0.290	-0.725	-1.945	**-2.005***	**-2.755****
	(1.606)	(1.039)	(0.862)	(1.216)	**(1.208)**	**(1.307)**
Sec1	**8.809*****	3.442	2.315	1.073	0.010	-2.543
	(1.829)	(2.348)	(1.721)	(2.403)	(2.469)	(2.553)
Sec2	1.910	**8.160*****	3.105	3.303	-0.840	-1.116
	(1.721)	**(2.678)**	(2.112)	(3.040)	(3.004)	(3.133)
Sec3	-1.340	3.016	-2.553	0.421	**-6.761****	**5.631***
	(1.585)	(2.470)	(1.912)	(2.699)	**(2.741)**	**(2.931)**
Age (2005)	-1.573	-0.462	-2.386	-0.606	-3.497	1.493
	(1.820)	(2.311)	(1.638)	(2.314)	(2.309)	(2.452)
Age squared (2005)	-0.0551	-1.574	-1.259	0.454	-1.628	-0.342
	(0.216)	(2.770)	(1.870)	(2.757)	(2.605)	(3.013)
Number of children (2005)	-0.890	0.022	0.102	0.175	0.312	0.086
	(2.025)	(0.313)	(0.223)	(0.312)	(0.280)	(0.338)
Married	-0.846	-2.549	0.891	-1.231	**-3.035****	1.922
	(0.826)	(2.837)	(2.158)	(1.224)	**(1.226)**	(1.288)
Selection_correction	-2.048	-1.288	-1.398	-2.804	-5.453	-1.492
	(2.920)	(1.201)	(0.866)	(4.198)	(4.121)	(4.299)
Attrition_correction	-3.546	-1.731	-3.707	**-8.051****	-2.454	-0.104
	(2.578)	(4.093)	(2.998)	**(3.659)**	(3.666)	(4.209)
State control variable	Yes	Yes	Yes	Yes	Yes	Yes
Constant	**31.07*****	**31.113****	**40.30*****	**38.990****	**54.442*****	-21.127
	(11.04)	**(15.354)**	**(11.43)**	**(15.855)**	**(15.286)**	(16.926)
Observations	6946	6946	6,946	6946	2386	4335
R-squared	0.173	0.162	0.130	0.139	0.048	0.065

Source: Author's calculations from the IHDS dataset
Notes: Household-level clustered standard errors (Equation 4 and 5) or bootstrapped standard errors (equation 6, 7 and 8; 500 replications) in parentheses; *** p<0.01, ** p<0.05, * p<0.1

The first three columns show that without implementing the instrumental variable method and/or without bootstrapping, gender appears to be a significant determinant of percentile change. Compared to men, women moved less across the hourly earnings distribution on an average ranging from 1.86 percentiles (OLS) to 2.30 percentiles (Bootstrapped OLS).

In all specifications, the coefficient of *initialY* is significant and negative. Estimation 7 shows that hourly earnings in 2005 are negatively associated with mobility. Moreover, compared to the OLS regression, the 2SLS estimator shows that the magnitude of the coefficient is significantly smaller when the endogeneity of initial earnings is controlled for. An increase of hourly earnings of 1% is associated with a decrease of 0.24 percentile change[65]. These results, therefore, imply that having a higher income is associated with less mobility across the distribution and that mobility is more likely for individuals who are at the bottom of the distribution.

After controlling for measurement error and for the endogeneity of initial earnings, gender is no longer a determinant of percentile change. Indeed, the coefficient is smaller and the standard error is higher, which suggests that measurement error may bias the coefficients of estimation 6. Regarding religion and caste groups, none of the groups have significantly different mobility patterns than the Hindu Upper Caste group regardless of the estimation method.

Estimations 8 and 9[66] respectively restrict the sample to individuals who experienced a downward percentile change (PC<0) and individuals who experienced and upward percentile change (PC>0). These estimations indicate that when other factors are controlled for, women experience more upward earnings mobility than men by 2.50 percentiles.

In the bootstrapped IV estimation, higher education and innate ability do not seem to be particularly associated with earnings mobility. Compared to individuals who have no formal education, having only primary education is negatively correlated to percentile change, and having a middle school education is positively correlated to percentile change. These correlations are both relatively small.

Finding adequate instruments when dealing with earnings and mobility can be challenging. Indeed, few variables can fit the criteria of theoretically being correlated with the level of earnings without being correlated to percentile change. Although the IHDS dataset is very rich, meeting these criteria is challenging. For these reasons, we adopted an empirical approach which consists in choosing the instrument that is theoretically *the least* correlated with percentile change while being empirically valid. Robustness tests of underidentification, weak identification and

[65]We estimated equation 3 by replacing initialY with its logged value and the results are very similar.

[66] Note that these two estimations are only presented for illustrative purposes since they were conducted on smaller samples and present a selection bias because of the sample restrictions.

overidentification show the empirical validity of the instruments. The instruments we use in the 2SLS estimations (an asset score and the average years of education at the PSU level) are robust to the aforementioned tests[67] as shown in Table 2.9.

Table 2.9. Instruments validity test

Test	Statistic	P-value
Underidentification test (*Ho: Model is underidentified*)	Kleibergen-Paap LM statistic = 578.029	0.000
Weak identification test	Cragg-Donald Wald F statistic = 382.410 (Kleibergen-Paap rk Wald F statistic) = 370.206 Stock-Yogo weak ID test critical values: 10% maximal IV size: 19.93 15% maximal IV size: 11.59 20% maximal IV size: 8.75 25% maximal IV size: 7.25	N.a.
Overidentification test (*Ho: Model is not overidentified*)	Hansen J statistic = 1.187	0.276

Source: Author's calculations from the IHDS dataset

Moreover, since our interest in this chapter does not solely lie in the *InitialY* variable, it could be argued that this variable should be excluded from the analysis altogether. Nevertheless, doing so would result in an omitted variable bias.

In the following sections, we provide two additional sets of estimations. In 6.2.2 we replace the dependent variable with a Weighted Percentile Change (i.e. PC reweighted by the logged earnings of 2005). By transferring the endogenous variable to the left-hand-side of the equation we estimate, we can compare the significance of the independent variables. In 6.2.3., we provide estimations of alternative quantile jumps (vintiles, deciles and quintiles) in order to verify the relevance of our bootstrapping method. Besides their purpose as robustness checks, these two additional sets of estimations also provide interesting additional information on the patterns of mobility.

6.2.2. Weighted percentile change estimations

Table 2.10 shows the results of bootstrapped OLS estimations which is not prone to the endogeneity of initial hourly earnings (InitialY) since the percentile variable is reweighted by the levels of log initial earnings. WPC gives more weight to movements of individuals with initially low incomes, which is why we provide estimation results on initial income ranges (by quintiles of initialY). The

[67] These tests were done on the non-bootstrapped equation as the bootstrapping does not allow to compute the necessary statistics for the tests.

results nuance the findings from our baseline estimations and shows heterogeneous effects depending on the ranks of initial income.

Table 2.10. Weighted Percentile Change estimations

	OLS estimations with selection correction				
	Weighted Percentile Change				
	Estimation 10	Estimation 11	Estimation 12	Estimation 13	Estimation 14
	If InitialY ϵ Quintile_1	If InitialY ϵ Quintile_2	If InitialY ϵ Quintile_3	If InitialY ϵ Quintile_4	If InitialY ϵ Quintile_5
Female	**-5.946*** **	**-6.366*** **	**-3.570*** **	-0.931	0.237
	(1.165)	**(0.911)**	**(1.097)**	(0.763)	(0.444)
Hindu OBC	-0.101	-0.869	**-1.420** **	**-1.036***	-0.142
	(1.723)	(1.022)	**(0.713)**	**(0.554)**	(0.399)
SCST	0.470	0.517	-0.951	-0.723	0.029
	(1.905)	(1.141)	(0.830)	(0.698)	(0.450)
Muslim Upper Caste	-0.153	-0.136	-1.538	**-2.581** **	-1.288
	(2.042)	(1.223)	(1.359)	**(1.224)**	(1.025)
Muslim OBC	-0.593	-0.757	-0.909	-0.871	-0.934
	(2.054)	(1.238)	(0.957)	(0.962)	(1.081)
Other	6.925	0.276	-0.541	-0.599	0.621
	(5.398)	(1.897)	(1.293)	(0.994)	(0.532)
Educ_primary (2005)	-0.614	0.105	0.348	0.380	0.410
	(0.644)	(0.414)	(0.328)	(0.286)	(0.251)
Educ_middle (2005)	0.007	-0.002	-0.005	-0.005	-0.005
	(0.008)	(0.005)	(0.004)	(0.003)	(0.003)
Educ_secondary (2005)	-0.571	0.263	**1.695** **	1.067	-1.092
	(1.666)	(0.882)	**(0.854)**	(1.025)	(1.552)
Educ_higher (2005)	-1.229	0.062	**2.078*** **	**1.633** **	0.591
	(1.225)	(0.627)	**(0.609)**	**(0.702)**	(1.155)
Sec1	**9.875***	2.373*	**2.849** **	**1.652***	**2.633** **
	(5.457)	(1.348)	**(1.399)**	**(0.978)**	**(1.117)**
Sec2	**11.877***	2.475	**6.531*** **	**2.437** **	**2.561** **
	(6.737)	(1.752)	**(1.687)**	**(1.099)**	**(1.157)**
Sec3	-3.398	1.057	-1.837	**1.963** **	0.620
	(7.084)	(2.199)	(1.736)	**(0.933)**	(0.491)
Age (2005)	-6.025	-0.564	-2.088	0.558	-0.059
	(6.066)	(1.526)	(1.504)	(0.839)	(0.470)
Age squared (2005)	**-10.636***	-0.012	-0.428	-0.202	0.604
	(6.264)	(1.810)	(1.615)	(1.031)	(0.633)
Number of children (2005)	**-2.809** **	-0.336	0.196	1.006	-0.542
	(1.146)	(0.741)	(0.711)	(0.649)	(0.563)
Married	0.092	0.033	-0.133	-0.074	0.138
	(0.325)	(0.196)	(0.183)	(0.168)	(0.143)

Table 2.10 continued on the next page

Table 2.10 (continued)					
Selection_correction	-3.585	-1.396	0.195	-2.153	**-5.481*** **
	(4.960)	(3.143)	(2.439)	(1.947)	**(1.400)**
Attrition_correction	-0.201	-0.557	0.418	-0.013	0.504
	(4.167)	(2.468)	(2.122)	(1.905)	(1.527)
Constant	**43.524** **	12.124	-3.881	-5.259	-7.548
	(18.741)	(11.806)	(9.367)	(7.719)	(5.999)
State control	Yes	Yes	Yes	Yes	Yes
Observations	1,491	1,481	1,411	1,344	1,219
R-squared	0.197	0.212	0.177	0.132	0.164

Source: Author's calculations from IHDS data
Note: Bootstrapped standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

The results concerning gender are consistent with estimation 7 for the quintiles 4 and 5, in which gender is not a determinant of earnings mobility. In the first three quintiles, being a woman is negatively associated with mobility. Concerning socio-religious groups, there is a significant negative association between being from the Hindu OBC group and earnings mobility (quintiles 4 and 5) and between being from a Muslim Upper Caste and earnings mobility (quintile 4). In almost all quintiles, the variables Sec1, Sec2 and Sec3 have a significant and positive coefficient, which indicates that innate ability is positively associated with earnings mobility. Sec2 has a higher coefficient than Sec1 in each quantile. Sec1 being the highest score, these trends suggest a non-linear association between the level of ability and earnings mobility.

Overall, the alternative specification shows that the effects we find in our bootstrapped 2SLS estimations are heterogeneous across the distribution.

6.2.3. Alternative quantile jumps

This aim of this section is twofold. Using alternative Quantile Changes (vintiles, deciles, quintiles)[68] as dependent variables, we can compare the results between bootstrapped and non-bootstrapped estimations to see if our method correctly corrects the measurement error issue. Furthermore, we can observe the determinants of these larger quantile jumps.

[68] QC_vintile refers to mobility across 20 earnings shares, QC_decile across 10 earnings shares and QC_quintile across five income shares. Therefore the possible values for each variable are [-20;20] for QC_vintile, [-10;10] for QC_deciles and [-5;5] for QC_quintiles.

Table 2.11. Alternative quantile jumps

	Alternative percentile jumps 2SLS estimations			Alternative percentile jumps Bootstrapped 2SLS estimations		
	Estimation 15	**Estimation 16**	**Estimation 17**	**Estimation 18**	**Estimation 19**	**Estimation 20**
	QC_vintile	**QC_decile**	**QC_quintile**	**QC_vintile**	**QC_decile**	**QC_quintile**
VARIABLES						
initialY	**-0.569****	**-0.309****	**-0.178*****	-0.470	**-0.331***	**-0.300*****
	(0.255)	**(0.127)**	**(0.064)**	(0.368)	**(0.182)**	**(0.101)**
Female	**-0.534*****	**-0.252*****	**-0.188*****	-0.150	-0.075	**-0.206****
	(0.190)	**(0.095)**	**(0.047)**	(0.275)	(0.132)	**(0.0801)**
Hindu OBC	**-0.287***	-0.128	**-0.073***	-0.218	-0.215	-0.058
	(0.173)	(0.086)	**(0.044)**	(0.250)	(0.132)	(0.0620)
SCST	-0.018	-0.006	-0.017	0.083	-0.049	-0.003
	(0.201)	(0.100)	(0.051)	(0.288)	(0.148)	(0.074)
Muslim Upper Caste	0.006	0.026	-0.004	-0.083	-0.207	-0.016
	(0.281)	(0.142)	(0.072)	(0.407)	(0.205)	(0.010)
Muslim OBC	-0.132	-0.065	-0.071	-0.145	0.002	-0.046
	(0.250)	(0.126)	(0.064)	(0.347)	(0.167)	(0.088)
Other	0.279	0.128	0.068	-0.021	**0.391***	0.188
	(0.312)	(0.157)	(0.080)	(0.466)	**(0.235)**	(0.123)
Educ_primary (2005)	0.005	-0.001	0.010	0.042	-0.025	0.032
	(0.237)	(0.118)	(0.060)	(0.313)	(0.170)	(0.086)
Educ_middle (2005)	-0.109	-0.031	0.009	**-0.534****	0.020	0.011
	(0.177)	(0.088)	(0.045)	**(0.240)**	(0.129)	(0.062)
Educ_secondary (2005)	0.479	0.252	**0.206****	0.221	0.113	0.062
	(0.344)	(0.173)	**(0.087)**	(0.506)	(0.258)	(0.128)
Educ_higher (2005)	0.525	0.292	**0.192***	0.446	0.043	0.155
	(0.410)	(0.206)	**(0.104)**	(0.612)	(0.300)	(0.160)
Sec1	**-0.641***	**-0.312***	**-0.201****	**-0.958***	-0.168	-0.003
	(0.361)	**(0.181)**	**(0.091)**	**(0.550)**	(0.274)	(0.131)
Sec2	-0.496	-0.247	**-0.169****	-0.682	0.076	-0.045
	(0.322)	(0.162)	**(0.082)**	(0.488)	(0.238)	(0.125)
Sec3	-0.225	-0.118	-0.081	-0.451	0.181	0.011
	(0.370)	(0.185)	(0.094)	(0.537)	(0.264)	(0.143)
Age (2005)	**-0.226*****	**-0.113*****	**-0.066*****	**-0.195***	**-0.248*****	**-0.071****
	(0.085)	**(0.042)**	**(0.021)**	**(0.115)**	**(0.0633)**	**(0.030)**
Age squared (2005)	**0.002****	**0.001****	**0.001*****	0.002	**0.003*****	**0.008****
	(0.001)	**(0.000)**	**(0.000)**	(0.001)	**(0.001)**	**(0.004)**
Number of children (2005)	0.023	0.010	0.001	0.100	0.006	-0.007
	(0.044)	(0.022)	(0.011)	(0.062)	(0.0309)	(0.016)
Married	-0.273	-0.119	-0.057	-0.367	0.040	-0.056
	(0.173)	(0.087)	(0.043)	(0.246)	(0.126)	(0.059)
Selection_correction	-0.883	-0.459	**-0.308****	-0.687	**-1.309*****	**-0.368***
	(0.596)	(0.298)	**(0.153)**	(0.871)	**(0.435)**	**(0.216)**

Table 2.11 continued on the next page

Table 2.11 (continued)						
Attrition_correction	-0.850	-0.475*	**-0.273****	-0.122	-0.141	-0.239
	(0.527)	(0.265)	**(0.133)**	(0.784)	(0.373)	(0.190)
State control variable	Yes	Yes	Yes	Yes	Yes	Yes
Constant	**9.305*****	**4.745*****	**2.750*****	**6.675****	**7.333*****	**3.078*****
	(2.263)	**(1.130)**	**(0.576)**	**(3.184)**	**(1.719)**	**(0.816)**
Observations	6,946	6,946	6,946	6,946	6,946	6,946
R-squared	0.141	0.142	0.145	0.138	0.160	0.182

Source: Author's calculations from the IHDS dataset
Note: Household-level clustered standard errors *** p<0.01, ** p<0.05, * p<0.1

Assuming that measurement error cannot be as high as to affect the allocation of an individual into the wrong vintile, decile or quintile, we do not need to apply the bootstrapping process to this step, which also gives it a purpose of providing a robustness verification of our bootstrapping methodology. For this purpose, equation 15, 16 and 17 are estimated using the 2SLS approach and equations 18, 19, and 20 are estimated with the bootstrapped 2SLS approach. Note that we interpret the first set of results and use the second set of results exclusively as a robustness test.

The bootstrap estimations for QC_decile and QC_quintile show signs and significance levels for the initialY variable that are similar to the non-bootstrapped estimations. For vintiles, the coefficient loses its significance with the bootstrapped method. It is possible that measurement error affects the distribution across vintile shares, consequently the loss of significance is not surprising. Nonetheless, the bootstrapped 2SLS results underestimate the earnings mobility differentials between male and female earnings for vintiles and deciles. The results concerning caste groups are mostly consistent with the 2SLS estimations except for the Hindu OBC group.

The estimations show that initial earnings are a significant and negative determinant of earnings mobility even for larger quantile jumps and that being a woman is negatively associated with the three types of mobility. There is no significant correlation between religion/caste and earnings mobility. The positive effect of education disappears for smaller percentile jumps (i.e. vintiles and deciles), which suggests that education only encourages larger percentile leaps (i.e. quintiles).

6.3. Correlations between both types of mobility

In order to observe how professional mobility is correlated to earnings mobility, we add each professional mobility variable in the 2SLS estimation of Percentile Change. The results are presented in Table 2.12.

Table 2.12. Correlations between occupational and earnings mobility

	PC-2SLS estimations		
VARIABLES	**Estimation 21**	**Estimation 22**	**Estimation 23**
Casual-Regular (base: downward mobility)			
No mobility	**5.235*** **		
	(1.854)		
Upward mobility	3.011		
	(2.268)		
Industrial mobility (base: no mobility)			
Mobility		0.319	
		(0.776)	
Skill mobility (base: downward mobility)			
No mobility			**3.892*** **
			(1.321)
Upward mobility			**6.581*** **
			(1.418)
InitialY	-0.194**	-0.178*	-0.161
	(0.0942)	(0.0921)	(0.0988)
Female	-0.549	-0.769	0.689
	(1.228)	(1.210)	(1.238)
Hindu OBC	-1.856	-0.331	-1.245
	(1.309)	(1.302)	(1.322)
SCST	-0.276	0.844	-0.578
	(1.460)	(1.455)	(1.517)
Muslim Upper Caste	-0.673	0.623	0.479
	(2.054)	(2.046)	(2.030)
Muslim OBC	-1.368	1.212	0.547
	(1.783)	(1.816)	(1.790)
Other	0.701	2.318	0.323
	(2.342)	(2.367)	(2.314)
Educ_primary (2005)	2.464	2.450	**-2.810***
	(1.709)	(1.659)	**(1.677)**
Educ_middle (2005)	0.394	1.070	0.0909
	(1.239)	(1.192)	(1.187)
Educ_secondary (2005)	1.517	**6.120***	3.504
	(2.559)	**(2.455)**	(2.449)

Table 2.12 continued on the next page

Table 2.12 (continued)			
Educ_higher (2005)	3.879	**7.396****	3.323
	(3.137)	**(3.037)**	(3.033)
Sec1	-0.0493	-4.372	-2.031
	(2.765)	(2.707)	(2.803)
Sec2	-0.740	**-5.089****	**-3.977***
	(2.437)	**(2.326)**	**(2.411)**
Sec3	1.415	-3.780	-2.543
	(2.737)	(2.802)	(2.746)
Age (2005)	-0.654	-0.982	**-1.324****
	(0.650)	(0.603)	**(0.628)**
Age squared (2005)	0.00627	0.00966	**0.0143***
	(0.00770)	(0.00712)	**(0.00740)**
Number of children (2005)	-0.0770	0.106	0.389
	(0.313)	(0.336)	(0.318)
Married	-1.103	-0.882	-1.806
	(1.262)	(1.208)	(1.313)
Selection_correction	0.658	-2.162	-3.952
	(4.510)	(4.277)	(4.391)
Attrition_correction	-9.303**	-3.987	-2.271
	(3.770)	(3.845)	(3.515)
State control	Yes	Yes	Yes
Constant	17.96	**30.12***	**34.55****
	(17.20)	**(16.59)**	**(16.67)**
Observations	6,789	6,947	6,947
R-squared	0.133	0.125	0.138

Source: Author's calculations from the IHDS dataset
Note: Bootstrapped standard errors in parenthesis (500 replications) *** p<0.01, ** p<0.05, * p<0.1

The results show that compared to downward mobility from regular occupations to casual occupations, no mobility is significantly and positively associated with earnings mobility but the coefficient for upward mobility is not significant. Surprisingly, casual to regular employment is not correlated to earnings. Industrial mobility is not significantly associated with earnings mobility. No skill mobility and upward skill mobility are both positively correlated to earnings mobility, with a stronger coefficient upward mobility than no mobility. Overall these results show that occupational mobility is not necessarily a vector of earnings mobility. The non-significance of industrial mobility also suggest that this type of mobility reflects employment change because of economic instability rather than to seize better earnings opportunities.

6. Discussion and conclusion

Adopting a labor market perspective on economic mobility indicate how the labor market contributes to increasing or decreasing horizontal inequalities. Table 2.13 and 2.14 summarizes the results concerning gender and socio-religious patterns of labor market mobility.

Table 2.13. Labor market mobility by gender

Gender	
Occupational mobility	
Description	*Inference*
Women are more immobile than men	No significant effects except for a positive association between being a woman and upward skill mobility
Earnings mobility	
Description	*Inference*
No significant differences	<u>Percentile Change</u> Significant negative association (OLS, Bootstrapped OLS, 2SLS) Non-significant association (Bootstrapped 2SLS) <u>Weighted Percentile change</u> Significant negative association for the first three quintiles <u>Other Quantile Changes</u> Significant negative association for vintile jumps (2SLS), decile and quintile jumps (2SLS, Bootstrapped 2SLS)

Source: Author

Table 2.14. Labor market mobility by socio-religious groups

Socio-religious groups	
Occupational mobility	
Description	Inference
Casual-to-regular mobility Higher upward mobility for Hindu Upper Castes. Highest downward mobility for Muslim OBCs Industrial mobility Hindu Upper Castes have important shares of stayers and transitions into the services sector Occupational skill mobility Limited upward skill mobility for SCSTs Higher skill mobility for Hindu Upper Castes	Casual-to-regular mobility Significant negative association between upward mobility of all groups compared to Hindu Upper Castes Industrial mobility Significant positive association between Industrial mobility and Muslim group Occupational skill mobility Negative association between upward mobility and SCST group Positive association between upward mobility and Muslim group
Earnings mobility	
Description	Inference
Process of catching-up with higher percentiles of mobility for all groups compared to Hindu Upper Castes	Percentile Change No significant associations Weighted Percentile Change Significant negative association for Hindu OBCs in the middle of the distribution (quintiles 3 and 4) Significant negative association for Muslim Upper Castes in the middle of the distribution (quintile 4) Alternative Quantile Jumps Non-significant effects except for a significant negative association between vintile and quintiles jumps for Hindu OBCs (2SLS)

Source: Author

The descriptive results show that religion and caste groups face different patterns of occupational mobility, often in the favor of Upper Caste Hindus. Depending on the type of mobility, the most disadvantaged group is not always the same (SCSTs in terms of low-skill immobility, Muslim OBC in terms of casual-regular occupational transitions). Industrial mobility, which cannot be ranked in terms of upward and downward mobility is a common occurrence for all caste groups and it is interesting to see that transitions into the services sector seems to be high for all groups, regardless of the industry they were initially in. The relative earnings mobility does not seem to differ significantly between men and woman. In terms of religion and caste, there is an indication of a process of catching-up since all groups seem to slightly benefit from more mobility than Hindu Upper Castes.

Besides the descriptive evidence, this study also proposes to estimate the determinants of occupational and hourly earnings mobility. The estimation of the determinants of labor market mobility shows that, when other factors are controlled for and when we apply several econometric corrections, gender does not remain a significant determinant of labor market mobility except for upward skill mobility. In the latter case, the coefficient of gender is positive which imply that the disadvantage women face in skill mobility is mostly due to important inequalities in education. Upward skill mobility being positively associated with earnings mobility, women are likely to benefit from this transmission channel. Nevertheless, the fact that women are less likely to experience vintiles, deciles and quintile jumps than men, this effect is small in magnitude. Both Muslim groups have a lower likelihood of experiencing casual to regular employment. Skill mobility is less likely for SCSTs and more likely for Hindu OBCs. Caste and religion are not directly associated to the level mobility in hourly earnings. Finally, compared to results that suggest absolute economic mobility in India, our analysis point to the fact that it does not necessarily contribute to erasing horizontal inequality between gender and socio-religious groups.

From a methodological perspective, this chapter proposes to correct many possible biases in the estimations. A contribution of our study is to generate the distributions of the hourly earnings mobility in a bootstrapping process which allows controlling for measurement error. A robustness test using higher quantile jumps to verify the results suggest that the methodology underestimates women's mobility patterns, most probably because there are fewer women in the sample and dropping some women from the sample during the bootstrapping process influences the results. This issue requires further investigations.

Furthermore, the values of the constants being high and the relatively small R^2 suggest that there are other determinants of percentile change that this study does not consider. Variations in social capital, for instance, could be the reason for unequal access to promotions in a given occupation. Nevertheless, in the case of India, the level of social capital differs significantly by castes (Munshi and Rosenzweig 2006), and part of its effect on mobility is therefore captured by the socio-religious variable.

Chapter 3. Heterogeneous patterns of earnings structure and segmented labor markets

1. Introduction

In developing countries, high rates of informal employment imply large shares of insufficiently remunerating jobs and a lack of social protection (de Laiglesia and Jütting 2009). The high prevalence of informal employment can hinder economic development and the reduction of poverty, because both do not only rely on job creation but also on specific employment characteristics such as the regularity of income.

The important economic policy shift that occurred in the 1980s in India, mainly consisting of extensive liberalization measures, has played an important role in the labor market. Prior to the liberalization, the post-independence economic policy encouraged the protection of import-competing industries through industrial regulation and a complex licensing process. In this context, opening the economy to competition and privatizing a large share of public industries had considerable impacts on the labor market (Chamarbagwala 2006). Indeed, there was an automatic decrease in public industrial employment. Moreover, the increase of competition from import and export trade liberalization led Indian firms to increase the demand for subcontracting and for casual labor on the one hand, and the decrease of formal employment combined with demographic changes created a large pool of available labor force for these activities on the other hand. These trends have contributed to the actual high shares of informal wage-workers and non-agricultural self-employment in India.

The Indian labor market has higher rates of informal employment than most developing countries, leading up to 82.2% of total employment in 2011-12 (ILO 2016a). In this context, considering the formal economy as a baseline and defining informality by comparing it to a residual share of employment can be misleading for the following reasons. First, the definition of the "*unorganized* *sector*" in the Indian legal framework is only an approximation of the "*informal sector*" as defined by the ILO because there is a large share of activities that can be unregistered, making it difficult to consider registration as a norm. Likewise, labor laws in India provide an important leeway for casual labor and contract labor which are closer to the ILO's definition of informal than formal labor, thus making it difficult to use a legalist criterion to define informality. As for many countries, Unni and Naik (2013) have shown that it is more relevant to interpret the formal-informal divide as a continuum rather than making a binary distinction in the case of India. Therefore, understanding how the heterogeneous nature of this labor market might translate into the coexistence of different segments that are more or less porous is challenging. Nevertheless, studying the existence and mechanisms of labor market segmentation is necessary for the right orientation of public policy towards vulnerable groups. For instance, if a given group is blocked in a poorly remunerating segment, the main issue will then be to increase intersegment mobility rather than focusing on educational policies. Informal sector employment comprises highest shares of working women than men globally. In India, women are mostly engaged in low productivity employment and are generally less paid than men, especially in the lower skilled manufacturing sector (Sorsa 2015; The World Bank 2012). Furthermore, the 61st round of the National Sample Survey (2004-05) shows that in urban areas, more than 40% of SCSTs engage in casual labor. As for Muslim workers, they rely more on self-employment than the other socio-religious groups (60% of the group is self-employed) and have less access to regular employment, especially for women (Government of India 2006).

The multifaceted reality of informal labor has been a topic of interest for economists for decades. Indeed, since the early work of Keith Hart in the 1970s (Hart 1973), economic research on the informal sector has had an important symbolic role in either challenging or consolidating different economic theories. On the one hand, economists have tried to analyze informality using existing frameworks to better understand its ins and outs for developing countries. On the other hand, the complex mechanisms observed in the informal economy have encouraged either the development of new frameworks or the adaptation of existing ones. One of the concepts that is often associated with the informal economy is *labor market segmentation*. Indeed, using this concept as a tool to understand the mechanisms of informal labor market arrangements has been largely recognized in the literature (Gindling 1991; Dickens and Lang 1985). Generally, the expression "labor market segmentation" refers to the coexistence of different subsets of the labor market, each one being subject to a specific set of formal and informal rules and regulations. The boundaries of a segment are more or less rigid and inter-segment mobility can be more or less constrained. The ongoing debate concerning informality in developing countries opposes two conceptions of informality, and therefore two types of labor market segmentation. The first one is an informal economy entirely composed of workers who enter the informal sector for subsistence-related reasons and because of a lack of opportunities in the formal sector (Gindling 1991). This type of labor market contains two homogenous segments, a formal one and an informal one. In the second conception of segmentation, a heterogeneous informal sector is composed of two segments: *a lower tier* composed of individuals who seek a job for *subsistence* and *an upper tier* composed of individuals who *choose* to be in the informal sector because of the potential opportunity it represents in terms of absence of regulation (Fields 2007; Perry et al. 2010). Empirical studies have pointed out the heterogeneous nature of the informal sector in several countries such as Côte d'Ivoire (Günther and Launov 2012), Indonesia (Rothenberg et al. 2016), Mexico (Alcaraz, Chiquiar, and Salcedo 2015) or Turkey (Salem and Bensidoun 2012). In the case of India, these questions need to be adapted. Indeed, considering that the formal labor market is residual, it is very likely to find heterogeneity in the informal labor market. Moreover, the earnings structures of own-account workers and small household businesses are likely to differ from wage workers'.

Labor market segmentation is a complex concept, often difficult to operationalize in empirical investigations. Although the multidimensional aspect of segmentation is evident in a theoretical perspective (Leontaridi 1998), parametric methods used for its detection require an *a priori* determination of segment-membership that tends to rely on a single criterion. In this case, it is necessary to define a segment based on a specific characteristic. When analyzing informality, criteria such as having an employment contract or being registered for production units can be used to distinguish formal from informal segments. However, using this legalist approach may not be the most relevant in the Indian case. Indeed, being unregistered is legal for all small production units and many alternative employment forms, namely contract labor, lay between formal and informal employment. Moreover, even if we established a list of criteria, our analysis would suffer from data limitation. Given the difficulty to establish such a list to distinguish the different tiers of the informal sector, semi-parametric methods are considerably helpful for the design of empirical models. By considering segment membership as a *latent variable* in the informal economy in Côte d'Ivoire, Gunther and Launov (2012) implemented an interesting approach using Finite Mixture

Modelling. Building on this study, the purpose of this chapter is to explore the heterogeneity of the urban Indian labor market. We consider two different "sectors" of the labor market: household businesses (which include own-account workers that may or may not benefit from contributing household members and more formal small enterprises) and salaried work. The separation of both sectors is motivated by the difference in the way income is calculated. We analyze the potential heterogeneity of both sectors by examining how many types of *earnings structure* can be identified in each of them. Indeed, a differential in the way productivity-related characteristics are linked to earnings in the Mincer-type earnings function (which we address as the "earnings structure") reflects the existence of different segments (Gindling 1991; Günther and Launov 2012). The results allow us to address the following questions: (i) Can we identify a duality that could relate to the formal *versus* informal duality among household businesses and among salaried workers? (ii) Is there an opportunity *versus* necessity form of duality in this predominantly informal labor market? (iii) How do social identity variables, namely gender, caste and religion, come at play in the segmentation process?

After accounting for the selection bias due to the non-random allocation of individuals into different types of employment, the study suggests that there is a homogenous household business sector and a segmented employment sector. Overall, belonging to specific socio-religious plays an important role in the way workers are sorted into different segments. Women do not have access to specific segments of the labor market and are grouped into one specific segment, suggesting employment segregation and possibly employment discrimination. The socio-religious stratification of the labor market is more nuanced but a concentration of the disadvantaged group exists in the lower segment of the labor market.

2. Informality in developing countries: concepts and literature

After defining informality this section reviews the literature on how the concept of labor market segmentation became a tool to understand the nature and the role of the informal economy in developing countries.

2.1.Definitions of informality

There are two main approaches to defining informality[69]: the *enterprise-based approach* (or employment-size approach) and the *labor approach*. The latter approach focuses on characteristics of individuals involved in the informal sector and allows a distinction between the informal sector and informal employment (Hussmanns 2004; Kanbur 2009). The "Resolution concerning statistics of employment in the informal sector" from the 15th International Conference of Labour Statisticians (ICLS) (1993) and the Guidelines concerning a statistical definition of informal employment adopted by the 17th ICLS (2003) lay the foundation for differentiating the informal sector, in which production units are considered as observation units, from informal employment in which jobs are considered as observation units. The contribution of this approach, besides an obvious ethical one regarding decent work and job quality perspectives, is to prevent any inclusion or exclusion errors concerning formal workers in informal units and informal workers in formal units.

As shown in Figure 3.1, informal employment in informal productions units is represented by own-account workers, employers, employees and members of producers' cooperatives. It can also be present in the formal sector in the form of informal employment in formal production units. Own account-workers are considered as self-employed individuals[70].

[69] Illicit activities which imply selling an illicit product or service are out of the scope of this study, as it is the case in most of the literature. Indeed, these activities have little in common with informal but licit activities.

[70] According to the ILO (1993), self-employment is composed of "*Employers that engage on a continuous basis one or more persons to work for them as 'employee'. Own-account workers have the same authority over the economic unit as the 'employers', but do not engage 'employees' on a continuous basis. Members of producer cooperatives take part on equal footing with other members in determining the organization of production etc. Contributing family workers cannot be regarded as partners in the operation of the productive unit because of their degree of commitment to the operation of the unit, in terms of working time or other factors, is not at a level comparable to that of the head of the enterprise.*"

Figure 3.1. Definition of informal employment

Production units by type	Jobs by status in employment								
	Own-account workers		Employers		Contributing family workers	Employees		Members of producers' cooperatives	
	Informal	Formal	Informal	Formal	Informal	Informal	Formal	Informal	Formal
Formal sector enterprises					1	2			
Informal sector enterprises[(a)]	3		4		5	6	7	8	
Households[(b)]	9					10			

(a) As defined by the Fifteenth International Conference of Labour Statisticians (excluding households employing paid domestic workers).

(b) Households producing goods exclusively for their own final use and households employing paid domestic workers.

Note: Cells shaded in dark grey refer to jobs, which, by definition, do not exist in the type of production unit in question. Cells shaded in light grey refer to formal jobs. Un-shaded cells represent the various types of informal jobs.

Informal employment: Cells 1to 6 and 8 to 10.

Employment in the informal sector: Cells 3 to 8.

Informal employment outside the informal sector: Cells 1,2, 9 and 10.

Source: Guidelines concerning a statistical definition of informal employment (ILO 2003)

2.2.Labor market segmentation as a tool to understand informality

2.2.1. Definition of segmentation in labor economics

According to the standard economic theory, in a perfectly competitive labor market with a homogeneous group of workers and firms, there is a single equilibrium where wages are set at market-clearing prices (Borjas 2012). Human capital theorists have used this baseline model as a starting point to explain human capital returns to wages in the 1950s. This has been allowed by relaxing the hypothesis of homogeneous agents on the supply side. In this framework, wages are linearly determined by the level of investment in human capital, which can be defined as an amount of accumulated knowledge, as well as personal characteristics that are observed or unobserved and that determine a person's productivity (Becker 1962; Mincer 1974). In this branch of literature, wages between two equally productive workers can only differ for the reasons listed below.

1. *Discrimination* which is only supposed to lead to short-term differentials. It is either due to employer's taste or to statistical discrimination, but it is supposed to disappear in the long-term because it implies an inefficient allocation of labor.

2. *Compensating differentials* mechanisms may exist in the labor market, implying that two equally productive individuals can have different wages if the worker with the lower wage receives other types of non-pecuniary advantages such as better working conditions, amenities etc. (Altonji and Blank 1999).

3. The existence of *non-competing groups*. According to Cairnes (1878, cited by Dimou (2006)), non-competing groups do not compete for the same types of occupations in the labor market because of factors such as belonging to different social classes or psychological reasons. A four-category classification of non-competing groups is presented by Dimou (2006), composed of non-qualified workers, craftsmen and retailers, engineers and businessmen and, professions of Science and Arts. The coexistence of these groups implies that the labor market is not perfectly competitive.

2.2.2. The formal and informal sectors

In the 1970s, the growing awareness of the existence of informality in developing countries led to the assumption that labor markets might be heterogeneous not only on the supply side (e.g. because of different human capital endowments) but also on the demand side. Multiple theories and definitions of informality emerged to constitute the three opposing views of the dualist, the structuralist and the legalist schools of thought. The institutionalist approach is a fourth school of thought which also proposed an analysis of informality.

The dualist theory finds its roots in the work of Lewis (1954) and Harris and Todaro (1970). In this framework, the informal sector is a residual part of the labor market composed of the labor supply that remains unabsorbed by the formal sector because of demand insufficiency. This setup implies that growth and structural change are supposed to lead to a shrinking, and ultimately to the disappearance, of the informal sector. Reich, Gordon, and Edwards (1973) define labor market segmentation as sub-markets or sub-processes with different characteristics, behavioral rules and working conditions. A feature that characterizes a segment and differentiates it from a threshold that one needs to overcome to have better wages is the imperfect mobility of agents from one segment to another. The formal or modern economy has better wages and better job quality than the informal sector. Indeed, the latter is only considered as a transitory phase for individuals moving from rural to urban areas, and who are only at the beginning of their active life. However, they tend to remain longer in the informal part of the economy because of formal-sector characteristics (such as minimum wages or trade unions) that keep wages above a market-clearing level. Therefore, two individuals with the same human capital endowments can have different labor market outcomes if they are in different segments of the labor market. By contrast, the structuralist school of thought considers that the informal sector is composed of small production units and unregulated workers that are subordinated to larger capitalist firms. This school of thought stems from the studies of Moser (1978) and Castells and Portes (1989). According to these authors, growth and structural change will not lead to a decrease in the size of the informal sector since the modern capitalist sector reacts to modernization by introducing more flexible systems. This approach points out that economic development relies heavily on informal firms, namely in the form of subcontracting relations between modern capitalist firms and small informal production units. The third school is the legalist school developed by De Soto (1989) who considers that the informal sector is comprised of entrepreneurs who do not want to bear the costs of registrations. The legalist studies point out the potential voluntary nature of being part of the informal sector. Empirical investigations have consequently questioned the role of the informal economy and the relevance of assigning it a segmented nature (which implies that workers are trapped in the informal sector) or a competitive nature (which implies that workers choose the informal sector) (Magnac 1991; Gindling 1991).

The fourth school of thought best qualified as the institutional economics view of informality emerged in parallel to the previous theories. The authors have a somewhat different and more firm-based approach to segmentation. Indeed, Doeringer and Piore (1985) consider that the labor market is composed of firms that have each an internal labor market component and an external one. The internal labor market is more stable and yields higher wages whereas the external one, being prone to market pressure, yields lower wages. The institutional rules that regulate the internal labor market create situations which cannot be explained using a human capital theory framework (Leontaridi 1998).

In the 2000's, the focus shifted from the analysis of the informal sector to informal employment, mostly driven by the ILO's work. Indeed, the sole consideration of an informal sector neglects some aspects of the informal economy such as informal employment in formal production units.

2.2.2. The heterogeneity of the informal economy

Up until the 2000s, empirical analyses of the informal sector have made the *a priori* assumption about its homogeneous nature (Günther and Launov 2012). However, its potential internal heterogeneity is becoming the source of considerable interest (Guha-Khasnobis and Kanbur 2006; Fields 2007). This branch of the literature considers the coexistence of two types of informal workers opposing those who are informal by *choice* to those who are informal by *necessity*. The lower tier of the informal labor market is composed of individuals who are involuntarily informal and the upper tier is composed of individuals who are in the informal labor market by choice (Fields 1994; Maloney 2004). This duality of the informal labor market generated academic debate because of its important policy implications, for instance in fiscal terms. The informal economy is originally depicted as the economy of the poor, which is why the potential non-precariousness of (at least a part of) individuals in the informal labor market can alter the objectives of labor market policies.

2.3. The specificities of informality in India

The literature on the universal criteria of the informal sector and informal activities provide a baseline to understand local specificities. Nevertheless, when analyzing a specific labor market, it is important to understand how formality, informality and their relationship are shaped by local formal and informal institutions.

A country's legal framework is a rich source of information on how informal activities and employment are perceived by institutions. In the case of India, although legal enforcement is a general concern, the institutional definitions of informality automatically excludes a large share of units from the necessity of registration and a large share of workers from the realm of labor law.

In India, the term informal is hardly used in the legal and regulatory framework. Instead, the closely related term of "*unorganized sector"* can be found. It identifies activities that are not in the *organized* economy. Kulshreshtha (2011) considers that this concept is an "approximation of the informal sector" as defined by the ILO[71]. Most of the definitions adopt an approach based on production units and distinguish an organized from an unorganized sector. The production units that constitute the unorganized sector are characterized by criteria closely related to the ILO guidelines such as a low level of organization, little or no division between labor and capital, and the absence of formal contracts. The National Commission for Enterprises in the Unorganized Sector (NCEUS) provides the most operationalizable definition of the unorganized sector, which is used by most statistical institutions today. The discriminatory identification criterion is the production unit size: enterprises with less than 10 workers (and less than 20 if the production unit does not have electricity) are qualified as unorganized. If these enterprises have more than the previously cited number of workers they are supposed to register under the Annual Survey of Industries (ASI). Otherwise, these units are also considered as unorganized. Although the ILO proposes to use the size of a production unit as a criterion to distinguish formal from informal units, neglecting the other criteria (such as the level of organization) can lead to a very heterogeneous group, that does not necessarily reflect informality. The choice of the NCEUS automatically includes a major part of economic units in the unorganized sector, regardless of their level of organization or capital. The implication of this categorization is that registration is not required for most activities and no distinction is possible between small organized production units and small labor-intensive unorganized ones. Moreover, since permanent workers from organized firms (which are registered under the ASI) are the only ones legally covered by the employment laws (Fagernäs 2010), the main part of the workforce is automatically outside of the realm of these laws.

According to the NCEUS, "*unorganized workers consist of those working in the unorganized sector or households, excluding regular workers with social security benefits provided by the employers and the workers in the formal sector without any employment and social security benefits provided by the employers"* (Sengupta et al. 2009). Informal employment is intrinsically linked to household businesses. In these production units, the contribution of each household member is often unclear and misreported in surveys. The threshold for being registered under the ASI can therefore be exceeded if, for instance, a small business owner (consciously or unconsciously) neglects reporting the contribution of household members to the activity.

---

[71] Appendix 3.1 lists the main institutional and legal definitions of informality in India based on the review of (Lee et al. 2008).

The share of workers in the unorganized sector was about 82.2% in 2011-12 and the share of informal employment was approximately 92%. The difference between both shares is equal to informal employment inside of the formal sector, taking the form of casual labor, which has significantly increased since 2004-05 (ILO 2016a). Casual labor often takes the form of contract labor (i.e. a contractor or a staffing firm links employers to workers). Contract labor remains on the boundaries of legality and is more likely to be considered informal than formal if we consider the ILO's definition of informal employment. There is therefore and important overlap between contract labor and casual labor in the Indian labor market.

Box 3.1. Contract labor in Ranipet

In a recent study, (Bertrand, Hsieh, and Tsivanidis 2017)[72] show that the rigidity of the Indian Disputes Act (IDA) has led employers to turn to contract labor more extensively in the last decades, especially after the liberalization of the economy which required firms to remain competitive. The authors find that this trend has been beneficial for growth but they do not consider the stakes of the informal nature of this type of labor. The authors state the following "*While staffing companies themselves have to abide by the IDA (like all formal firms), the contract workers they place into their customer firms are not formally employees of the customers.*". They do not address the fact that the Indian regulation does not require small staffing companies to register (if they have less than 10 or less than 20 employees and no electricity) and that there is little or no control for these contractors to register even if they have more than the legal threshold of employees. Therefore, there are no guarantees that all casual workers (if any) benefit from social security.

Contract work appears as a common form of employment in Ranipet. The terms of "*contractor*" or "*contract worker*" were mentioned regularly and promptly during the discussions on employment. Most interviewed individuals have a connection to this form of labor, either in their personal professional experience, or in their relatives'. In the same production units (tanneries) both contract workers and regular workers (temporary or permanent) can be employed.

In the specific context of Ranipet, there is no evidence of hiring firms or contractors being members of "*staffing companies*." Instead, they operate on their own (without registration) and the workers they hire do not benefit from any form of social security (Employee's Provident Fund and

[72] Note that this is paper in unpublished.

Employee State Insurance). Contract labor is described as a particularly precarious and unstable. For instance, although few women work in leather tanneries, there are some tasks such as leather drying that are operated by them. In this case, tanneries use contract labor and the demand fluctuates considerably depending on the weather, more specifically on the absence of rain. Consequently, these women face unemployment for relatively long periods if they are unable to find other employment opportunities.

For employers, contract labor is a very interesting option because it allows a short-term adjustment of the amount of labor according to the exact demand. Lower costs also motivate employers to use contract labor, the only costs being the wages (which are lower than for regular employees) and the commission paid to the contractor.

Some of the workers have this form of labor during their entire active life. For others, engaging in contract labor is the easiest way to enter the labor market, especially when they have a smaller social network. It is also a way of re-entering the market after an interruption (e.g. after a pregnancy for women). Contract labor is also a form of secondary activity that workers fall back on when there is a punctual need for extra income, which is especially visible for men. Compared to other type of jobs, contract labor seems easier to find than other forms of labor. Moreover, according to one of the workers, contract labor is more frequently found than "*regular jobs*" since a few years. This might be caused by a more fluctuant or unstable demand for leather goods in the case of Ranipet. Contract labor, therefore, echoes the structuralist theory of informality as the development of the town relies on this type of labor. The institutional view of informality, with an internal and external labor market responding to very different rules inside of the production units, also describe the fact that in a given firm it possible to have to very different forms of career-paths: regular employment and contract labor.

Source: Author

3. A methodology to analyze informality and labor market segmentation in India

This section proposes an empirical analysis of labor market segmentation in the context of urban India. The main research question of this study is to establish the homogeneity or heterogeneity of the labor market structure. We also analyze the specific roles of gender, religion and caste in the process of labor market segmentation. Finally, we discuss whether the segments of the labor market translate into a *formal versus informal* duality or a *necessity versus opportunity* one. A practical definition of a segment that will constitute the base of our analysis is to consider it as a part of the labor market that has a specific earnings function. Different earnings functions in the labor market imply that there are different returns to personal characteristics such as education, which crystallize the segment-specific mechanisms.

We implement a Finite Mixture Model (FMM) to simultaneously estimate the number of segments and the earnings function of each segment of the labor market. This method aims at detecting whether the density function of a variable (hourly earnings in our case) is the mixture of several density functions, each belonging to a specific latent segment. FMMs has previously been used to analyze the heterogeneity of the informal sector without having to establish the criteria of the informal segments beforehand (Günther and Launov 2012; Battisti 2013). However, in these studies, the authors distinguish formal from informal workers and exclude the formal segment from the estimation. Although our research question is close to ones in the previously cited studies, we implement a different methodology. Indeed, this study proposes an alternative *relative partition* for the following reasons. In the case of India, the distinction of formal and informal sectors or formal and informal employment can lead to very heterogeneous groups. As shown in section 2.3, if we follow the Indian guidelines related to the definition of informality, the formal sector a residual part of the economy. Moreover, as our focus lies on workers, it is very difficult to distinguish formal from informal employment whether we follow the Indian guidelines or the ILO ones. It is possible to distinguish salaried workers from casual workers. However, this is likely to result in groups that do not reflect a form of formal/informal distinction. These arguments motivate pooling all samples into a single simultaneous estimation of latent classes and earnings functions. Nevertheless, earnings from businesses and salaried occupations cannot be estimated at the same time since the nature of earnings is different between both groups. To compute individual business income from the IHDS dataset, we would have to divide the households' business income by the number of hours worked for each member which is less relevant than analyzing the household's income as a whole. Therefore, we will conduct a separate analysis of the potential heterogeneity of the household business sector and the salaried sector.

Considering segment membership as a latent multinomial variable (i.e. a given segment of the labor market), we estimate a set of Mincer-type income functions using a Finite Mixture Model. This method allows us not to make too many assumptions on what characterizes a segment in terms of outcome (for instance in casual labor or in regular labor). Indeed, predetermining the number of segments and what is considered to be a given segment of the labor market can lead to substantial bias (Dickens and Lang 1985). As in Gunther & Launov (2012) "*we let the data determine the number and size of informal segments that best describe the data*". The trade-off is that an important role is given to the variables that constitute the earnings function since the differences in the earnings structure will determine the existence of the segments.

3.1.Correcting the sample selection bias

Since the analysis is based on declared earnings, there is a selection bias linked to the fact that the sample of individuals who are active in the labor market and who have non-missing earnings information is not a random subsample of all working-aged individuals[73]. As shown in Chapter 2, this bias can be corrected using a Heckman two-step selection correction method.

Consider the following earnings function:

$$lny_i = x_{1i}\beta_1 + \varepsilon_i \quad \text{[Eq. 3.1]}$$

lny_i is the natural logarithm of hourly wages, x_{1i} is a vector of personal characteristics which determines wages and u_i is the error term. In this regression, the selection bias exists by construction since the regression only considers positive values of y_i. This implies that the sample of interest (labor market participants with non-missing earnings) is not necessarily a random subsample of the whole population. This bias causes β_1 to be inconsistent. The correction method

[73] A general reference for this section is Cameron & Trivedi (2005)

consists in augmenting equation (1) with the Inverse Mill's Ratio (IMR) computed from the probit selection equation (2).

$$\Pr(Part_i = 1) = \gamma(\beta_j x_2) \text{ [Eq. 3.2]}$$

$Part_i$ is a dummy variable which takes the value 0 if the person has no professional activity or missing income and the value 1 otherwise. x_2 is the vector containing explanatory variables for labor market participation. From this equation we can compute the inverse Mill's ratio (IMR) which is added as a regressor in Equation (1) such as:

$$lny_i = x_{1i}\beta_1 + \lambda(x_{2i}\widehat{\beta_2}) + \varepsilon_i \quad \text{[Eq. 3.3]}$$

Since this study contains two sets of estimations (household business earnings and salaried work earnings), we will need to correct the sample selection bias in both cases. In the first case, a Heckman correction method would consist in accounting for the fact that all households do not engage in business work. As for salaried work, this method would control for the fact that all working age individuals do not have a salaried occupation. In both cases, it can be argued that a simple binary selection between the selected sample and the rest of the population is insufficient. A more robust selection correction method would require a finer distinction of the main household activity, in order to distinguish households that engage in business work, in salaried work and that have different or no income sources. In the case of salaried workers, they must be differentiated from working-age individuals who are unemployed or inactive and business workers. In both cases, the sample selection bias is better specified as a multinomial logistic model. We use the implementation of the Durbin-McFadden selection correction method proposed by method (Bourguignon, Fournier, and Gurgand 2007) in order to estimate a multinomial selection equation and generate consistent correction terms. Using Monte Carlo simulations, the authors show that this method allows correcting the sample selection bias even if the independence from irrelevant alternatives (IIA) hypothesis is violated. Similarly, to the Heckman selection method, the estimation requires two steps. A multinomial logistic regression first estimates the probability of being in the salaried employment sector, in the household business sector or being inactive. Three correction terms which refer to the number of categories in the selection equation are then added to the earnings function.

As suggested in the literature (Nordman and Roubaud 2009; Dimova, Nordman, and Roubaud 2010; Ben Yahmed 2018), this method requires making exclusion restriction assumptions based on variables such as the number of infants, the number of children and the number of elderly individuals in the household.

Unlike previous studies (Günther and Launov 2012; Battisti 2013) that have implemented the Heckman method, our study is the first, to the extent of our knowledge, to consider the multinomial nature of the sample selection bias in a segmentation analysis using a Finite Mixture Model framework.

3.2. Finite mixture of regressions

In a Finite Mixture Model, the segment that a worker belongs to is considered as a latent variable. The framework described below estimates the probability of segment membership as well as each segment's earnings function simultaneously.

3.2.1. Estimation method

If we assume that the individuals from our sample are drawn from a population containing a discrete number of latent classes, the information for individual class membership can be considered as a missing variable. Finite Mixture Models are used to model distributions whose true density is a mixture of several density functions. This type of distribution occurs when the population is composed of different sub-populations that do not have the same density functions. In labor market economics, this method was initially used by Günther & Launov (2012) for the analysis of the informal sector in Côte d'Ivoire. The FMM shows whether the whole population has one Mincer-type earnings function or if there are different sub-segments, each having a specific earnings function. One of the most interesting advantages of this method is that the segments are not determined beforehand. Indeed, the FMM is a semi-parametric model in which the information regarding the segment each individual belongs to is unknown.

$$f(y_i|S;\ \theta_s) = (2\pi\sigma_s^2)^{-\frac{1}{2}} exp - \frac{[yi - \sum_j \beta_{j,s} x_{i,j}]^2}{2\sigma_s^2} \quad \text{[Eq. 3.4]}$$

If we consider that the earnings regressions follow a normal distribution, the normal density function of a worker's earnings[74] y_i conditional to belonging to a segment S is given by expression 4. β_s and σ_s are the regression coefficients and standard deviations respectively, while x_i are the regressors of the Mincer-type earnings function (Battisti, 2013). The parameters of the mixture density are estimated through an Expectation Maximization Algorithm.

Finite mixture models can be extended in order to include variables that predict class membership. The probability of a latent class (S) membership can either be considered as random or estimated using a multinomial logistic model where $P(S = |z_i)$. z_i, the latent-class predictors, are variables that predict class-membership. They can overlap with *xi* (Muthén and Asparouhov 2009; Wedel 2002). Lamont, Vermunt, and Van Horn (2016) enumerate two rationales for adding predictor variables in FMMs. First, it relaxes the implicit assumption that class membership and x_i are unrelated. Second, if there is a theoretical reason to believe that a set of z_i variables can predict class membership, the model will yield more precise estimates if they are added as predictor variables. However, the limitation in the use of predictor variables is that only a few variables can be added in order to allow the model to converge. This feature proves interesting in our case since gender, religion and caste are potentially linked to class membership. As pointed out by Doeringer and Piore (1985), segments of the labor market need to reflect class, racial and gender stratification of the society. Indeed, adding these variables allows to consider the different decision patterns between male and female workers as well as affirmative action policies and self-selection into occupations that may cause individuals from different socio-religious groups to allocate differently into the segments of the labor market. Only if these factors are considered in the allocation of workers into different segments, can we see the returns to endowments in each segment.

3.2.2. Optimal partition of the labor market

We estimate the FMM with different partitions of the labor market and then compare the estimations with a criterion that allows us to establish whether there is a continuous household business sector or whether it is segmented. We repeat the operation for the salaried employment sector. As suggested in the literature, we use Information Criteria as a comparison tool to gauge

[74] The same demonstration can be made for household business earnings by replacing the individuals regressors of the Mincer earnings function by determinants of business earnings at the household level.

the quality of the model and to choose the one that corresponds the most to the data (Günther and Launov, 2012; Battisti, 2013). Cameron and Trivedi (2005) advise using a criterion that penalizes the number of parameters the most, which is a property of the Bayesian Information Criterion (BIC).

3.2.3. Composition of each segment

Once we know the mixture of densities we can calculate an individuals' probability of being in each segment.[75] As in Günther & Launov (2012), we can assign individuals to segments based on the highest probability of membership, in order to describe each segment. Note that this probabilistic method does not aim at predicting exactly which worker will end up in which segment, but it provides information on the characteristics that influence segment-allocation. We can also analyze the characteristics of each segment by using variables that were not included in the estimation process such as the variables related to employment outcomes.

3.2.4. Detecting a formal/informal divide or a necessity/opportunity divide

Our methodology can be used to provide insights on whether there is an opportunity versus necessity segmentation (Günther & Launov 2012) but also on whether the earnings function that structure the labor market reflect a formal versus informal duality. Since we consider that having a specific earnings function defines a segment (and not criteria chosen beforehand), the FMM estimation results allow us to discuss both types of divides. Furthermore, including socio-religious group characteristics as predictor variables allows us to see their roles in the segmentation process.

Using the characteristics of individuals who are most likely to be in each segment, we can observe whether their labor market situation (e.g. labor market conditions, type of employment, skills, pay rate, etc.) reflect a formal versus informal segmentation. Moreover, based on their characteristics it is possible to estimate the potential earnings of each worker if they were assigned to the other segments. Comparing the different predicted earnings reveals potential low-earnings traps.

---

[75] The expression of posterior probability is : $p_{i,s} = \frac{\widehat{\mu_s} f_{i,s}(y_i | x_{i,j=1,j} \hat{\sigma}_s^2 \hat{\beta}_{j=1,-j,s})}{\sum_{s=1}^{k} \widehat{\mu_s} f_{i,s}\left(y_i \middle| x_{i,j=1,j} \hat{\sigma}_s^2 \hat{\beta}_{j=1,-j,s}\right)}$

4. Data description

4.1. Dataset

We use the 2011-12 wave of the IHDS dataset, excluding rural areas in order to focus on the urban dynamics of segmentation. The study is conducted in two separate steps. A first estimation concerns business earnings at the household level. The number of urban households is 10,102, 3,012 of which engage in business work as their main income source, while 1,974 have non-missing income or working hours values. The second estimation concerns salaried workers. The sample of all working-aged individuals (between 15 and 65, both values included) who are not currently students is composed of 40,525 individuals among which 14,175 have a salaried occupation as their main professional activity and do not have missing earnings. Extreme values (below the first and above the 99th percentile) were excluded for both types of earnings.

4.2. Model specifications and variable description

This section presents the specifications and the description of variables used for the selection equation and the household or individual earnings equations.

4.2.1. Selection equations

Since the samples used in the Mincer functions are not representative of the entire working-age population, correcting the sample selection bias is necessary for the robustness of the results. Our estimations of interest are at the household level for the business sample and at the individual level for the salaried workers sample. In both cases, we correct the sample selection bias by estimating a multinomial selection equation and computing selection terms that will be included in the earnings functions using the method developed by Bourguignon, Fournier and Gurgand (2007)[76]. We estimate both selection equations using the following exclusion variables: number of male and female infants (younger than five years old), number of male and female elderly individuals in the household (older than 65 years old). We believe these variables to be good predictors of labor market selection without affecting the level of earnings. The correlation tables in Appendix 3.2

[76] This method is implemented using the *selmlog* command in Stata.

show that in both cases, these variables are very mildly correlated to business earnings with correlation coefficients between 3% and 9% respectively, and to salaries (between 0% and 6% respectively). We also include the rank of individuals in the Secondary School Leaving Certificate examination which ranks from 1 to 3 (1 being the highest rank and individuals who do not pass the examination being coded 0). In the case of the household selection equation, we include the household average rank among adults. In the case of the individual salaried earnings function, we include a dummy that indicates if the person was ranked in the First class or not. A binary variable is also added to specify if a credit was contracted for the purposes of a business household selection equation (coded 0 or 1).

Table 3.1. Variables used in the selection equations

Selection equation at the household level	***Selection equation at the individual level***
Dependent variable	
Labor market participation (household)	Labor market participation (individual)
Exclusion restrictions	
Number of female infants (<5 y.o.)	
Number of male infants (<5 y.o.)	
Number of elderly female household members	
Number of elderly male household members	
Independent variables	
Religion/ Caste	Religion/ Caste
Household male head age	
Household male head age (squared)	Age
Household female head age	Age (squared)
Household female head age (squared)	
Highest education level (number of male household members)	
Primary	
Middle	
Secondary	Highest education level
Tertiary	Primary
Highest education level (number of female household members)	Middle
	Secondary
Primary	Tertiary
Middle	
Secondary	
Tertiary	
Average secondary class	Secondary class (base 0)
Credit	
Number of male children (between 5 and 15 y.o.)	Number of male children (between 5 and 15 y.o.)
Number of female children (between 5 and 15 y.o.)	Number of female children (between 5 and 15 y.o.)
Number of household members	Number of household members
State	State

Source: Author

4.2.2. Earnings function specification for household business workers

The IHDS dataset does not allow to correctly construct a common earnings variable for the self-employed and salaried workers. This is our main motivation for separating the two samples in this study. In the case of business workers, the dataset contains information concerning the net profit for household businesses at the household level, but information concerning hours worked is available at the individual level. We construct an average hourly earnings variable at the household level by dividing the average net profit of the business by the number of hours worked. The estimation is then conducted at the household level.[77]

The variables used in the specification of the Mincer-type earnings functions are shown in Table 3.2. The independent variables are the predictors of earnings, household characteristics and State control variables. Since earnings are estimated at the household level, we need to include appropriate education and productive ability variables. In order to avoid a high autocorrelation between educational characteristics of each household member, we choose to include the number of years of education of the male household head and of the female household head. Moreover, we compute the Secondary School Leaving Certificate average rank as an indicator of innate ability. Since the estimation concerns business earnings, we add a dummy variable that indicates whether individuals contracted a credit for the purposes of the business.

The statistics in Table 3.2 show that there are very few female decision makers in the sample (1.57%). Hindu Upper Castes and OBCs are highly represented (30.09% and 31.91% respectively). Muslims are also highly represented. Indeed, there are 2.60% of Muslim Upper Caste households and 6.93 % of Muslim OBC ones in the whole sample whereas their share in the estimation sample is 8.30% and 13.22% respectively. Conversely, although 20.35% of households are SCSTs, only 12.71% engage in business work. In a descriptive analysis using Census Data, Iyer, Khanna, and Varshney (2011) show that OBCs have made considerable progress in terms of entry into business work between 1990 and 2005 whereas SCSTs remain highly underrepresented and have smaller enterprises than the other groups. Their main hypothesis for this difference is that SCSTs suffer from having smaller networks which impedes their potential for starting or growing a business. Concerning hourly income, significant gaps are visible depending on the whether the decision-

[77] Note that when households reported having more than one business, we only considered the main one.

maker is a woman or on the socio-religious group.[78] Hourly wages are significantly higher for Hindu Upper Castes (INR 45.64) compared to the other groups. SCSTs have the lowest hourly income (INR 28.88), followed by Muslim OBCs (INR 29.73).

Table 3.2. Variables of the household business earnings function

Dependent variable	**Mean or percent**	**Standard deviation**	**Average hourly net profit (logged)**
Average hourly net profit (logged)	3.614	1.020	
Independent variables			
Female decision maker	1.57%	-	3.272
Male decision maker	98.43%		3.605
Religion/ Caste		-	
Hindu Upper Castes	30.09%		3.821
Hindu OBC	31.91%	-	3.529
SCST	12.71%	-	3.363
Muslim Upper Caste	8.30%	-	3.520
Muslim OBC	13.22%	-	3.392
Other	3.75%	-	3.986
Household male head age	46.009	9.380	
Household male head age (squared)	2206.655	869.855	
Household female head age	40.899	9.469	
Household female head age (squared)	1761.525	789.553	
Years of education (male household head)	9.685	4.475	
Years of education (female household head)	8.417	5.107	
Average SSC class	0.345	0.381	
Credit	20.26%	-	
Number of male children (between 5 and 15 y.o.)	0.751	0.890	
Number of female children (between 5 and 15 y.o.)	0.639	0.906	
State	-	-	

Source: Author's calculations from IHDS data

4.2.3. Earnings function specification for the salaried workers

An earnings function at the individual level includes determinants of earnings, personal characteristics and control variables. The type of variables that should be included in the latter group is debatable. Strictly supply-side variables should be included as they reflect either inherent characteristics of workers, or their choices such as investment in human capital. However, many potential control variables such as the type of occupation or the number of hours worked can also be demand-side variables as they can potentially indicate an outcome of employer practice rather than a productivity-related choice (Nordman and Roubaud 2009). We estimate hourly earnings so

[78] Appendix 3.3 shows the kernel density plots of hourly net profit

that the number of hours worked is considered as an outcome. We include the industry of the worker as a control variable since it does not necessarily reflect an outcome (as would the type of occupation for instance), but makes the earnings function more precise. Controlling for the sector of occupation or the industry is commonly observed in studies that use Mincer-type equations (see for instance Bargain and Kwenda (2011)) and in studies that analyze labor market segmentation with Finite Mixture Models such as the models of Günther and Launov (2012) and Battisti (2013).

We use age and its squared measure as a proxy for experience. It is also possible to calculate potential experience by subtracting the number of years of schooling and the age of entering school from observed age. However, in the case of India, this proxy might be biased because it is likely to overestimate the experience of the unschooled or poorly schooled workers (Goel 2017) and it is also prone to overestimating the experience of individuals who have discontinuous labor market experience which is particularly plausible for women (Nordman and Roubaud 2009).

In order to observe the returns to education, the highest education level is indicated with five dummy variables: no education (reference group), primary education, middle, secondary and tertiary education. A dummy variable for those who have an SSLC first-class level is included to control for innate ability.

The statistics in Table 3.3 show an overrepresentation of SCSTs and OBCs in salaried work which is logical provided that these groups are the most likely to work and that they are underrepresented in the self-employment and business sector. Indeed, although they represent 22.23% and 30.72% of the urban adult population respectively, their shares in the salaried employment population are 27.33% and 31.23% respectively.

Table 3.3. Variables of the salaried employment earnings function

Dependent variable	**Mean or percent**	**Standard deviation**
Hourly earnings	3.365	0.824
Independent variables		
Female	19.98%	-
Religion/ Caste		-
Hindu Upper Castes	24.56%	
Hindu OBC	31.23%	-
SCST	27.33%	-
Muslim Upper Caste	5.62%	-
Muslim OBC	7.78%	-
Other	3.46%	-
Age	37.821	
Age (squared)	1571.451	11.877
Education level		
None		
Primary	7.08%	
Middle	26.03%	
Secondary	14.78%	
Tertiary	30.51%	
SSC First Class	13.64%	
Sector of occupation		
Agriculture	5.25%	
Manufacturing	22.14%	
Services	52.83%	
Public Administration	6.88%	
Construction	12.90%	
Married	70.72%	
Number of female children (between 5 and 15 y.o.)	0.402	0.695
Number of male children (between 5 and 15 y.o.)	0.438	0.711
State	-	-

Source: Author's calculations from IHDS data

5. Results

The aim of our analysis is to determine if the structure of the labor market can be considered as "homogenous", that is a single earnings structure with the same returns to personal characteristics for all workers. In order to conduct the study, we separate the earnings function estimation for business workers and for salaried workers because we cannot merge both analyses due to of data limitations. In both sectors, we estimate an FMM model to detect latent variables.

The results are organized as follows: in Section 5.1, we provide evidence of a homogenous household business sector and in Section 5.2 we show that the salaried sector is segmented. Section 5.3 proposes a discussion on how our findings contribute to the formal *versus* informal, and to the opportunity *versus* necessity debates.

5.1. A homogenous business sector

This section presents the different results concerning the household business sector, leading to the conclusion that it constitutes a homogenous business sector.

5.1.1. Estimating the selection of households into business work

In order to apply Bourguignon, Fournier and Gurgand's (2007) selection correction method, we first estimate a selection equation with three possible categories of the main household income source: business work, salaried work and a third category including no income or other income source. The choice of the main source of income is necessary to classify households into different categories as in a given household there are different possible combinations of business work, salaried work and other income sources.

Table 3.4. Multinomial logistic estimation of the household main income source

	Main household income source (base category: salaried work)	
Variables	**Business work**	**Other**
Hindu OBC	**1.149****	1.086
	(0.074)	(0.131)
SCST	**0.502*****	0.869
	(0.038)	(0.118)
Muslim Upper Caste	**1.381*****	1.130
	(0.138)	(0.251)
Muslim OBC	**1.716*****	0.878
	(0.163)	(0.193)
Other	**1.362****	0.720
	(0.175)	(0.178)
Male head age	**1.151*****	**0.794*****
	(0.045)	**(0.060)**
Male head age (squared)	**0.999*****	**1.003*****
	(0.000)	**(0.001)**
Female head age	0.965	1.068
	(0.032)	(0.075)
Female head age (squared)	1.000	1.000
	(0.000)	(0.001)
Highest male education (in years)	1.007	**1.050*****
	(0.007)	**(0.014)**
Highest female education (in years)	**1.032*****	1.018
	(0.006)	(0.011)

Table 3.4 continued on the next page

Table 3.4 (continued)		
Average secondary class score of the household	0.892	1.272*
	(0.068)	(0.183)
Number of household members	**1.055*** **	**0.799*** **
	(0.019)	**(0.030)**
Number of female children (between 5 and 15 y.o.)	0.975	1.122
	(0.036)	(0.099)
Number of male children (between 5 and 15 y.o.)	1.015	**1.232** **
	(0.039)	**(0.108)**
Number of elderly male household members	**1.290*** **	**2.585*** **
	(0.111)	**(0.379)**
Number of elderly female household members	**1.191*** **	**1.331** **
	(0.077)	**(0.157)**
Number of female infants (<5 y.o.)	0.993	**1.432*** **
	(0.054)	**(0.155)**
Number of male infants (<5 y.o.)	1.075	1.113
	(0.059)	(0.130)
Constant	**0.021*** **	1.321
	(0.013)	(1.570)
Total Observations (3 categories)	10927	
Log likelihood	-40474.98	
LR chi2(100)	2477.47 (Prob > chi2 = 0)	

Source: Author's calculations from IHDS data
Note: Relative risk ratios (Exponentiated coefficients). Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.4 presents the estimation results in exponentiated form, they can therefore directly be interpreted as relative risk ratios. The results show that when all other factors are held constant, belonging to a specific religion or caste group plays an important role in determining whether a household engages in business work as its main income source, or if they engage in salaried work. SCSTs are 50.2% less likely to engage in business work than in salaried work compared to Upper caste Hindus. All other groups are more likely to have business work as their main income source compared to the same group. These results confirm the observation that SCSTs are underrepresented in the business sector, even when other factors that account for education and innate ability are considered. Moreover, a one-year increase in female education is associated with an increase in the probability of a household to engage in business work by 3.2%.

5.1.2. Estimation results

We explore the possible heterogeneity of the business sector by estimating the labor market partitions and the earnings function simultaneously with a Finite Mixture Model. We first present the statistics allowing us to choose the partition that best describes the data. We then present the household earnings structure of the business sector.

5.1.2.1. Partition and segment membership

The estimation procedure for the FMM consists in estimating several partitions and choosing the model that is a better fit of the data. Since it is recommended to use a criterion that highly penalizes the number of parameters (Cameron and Trivedi 2005; Battisti 2013) to choose the model, we use the BIC. The model with the smallest BIC is the one which contains the most information while staying parsimonious. Table 3. shows that in our case the single-segment OLS model is the most relevant one. The segmented model is not a better fit for the data. The choice of a single-segment model implies that in the household business sector, the earnings structure is the same for all households. There is a linear relationship between the estimated characteristics and household earnings and there is no existence of segments, or in other words of traps. As a consequence, if productivity-related characteristics have a positive and significant effect on household income, acquiring additional skills will allow the household to have a higher income. Nonetheless, the homogenous earnings structure does not necessarily confer a horizontal equality characteristic to the segment. There might indeed be group disadvantages, but they are not as high as to trap individuals in a low-earnings segment. Analyzing the earnings structure of this homogenous business sector will allow us to determine which household characteristics are rewarded and which ones are penalized.

Table 3.5. Bayesian Information Criterion of the business sector models

Model	Observations	Log-likelihood	Degrees of Freedom	Bayesian Information Criterion
1-segment (OLS)	1978	-2714.942	48	5794.196
2-segment	1978	-2575.288	96	5879.200
More than 3 segments	*Model does not converge within 150 iterations.*			

Source: Author's calculations from IHDS data

5.1.2.2. Earnings structure

The household business sector, which contains all self-employed individuals in the labor market is a homogeneous sector in terms of earnings structure. The earnings function that best characterizes this sector of the labor market is shown in Table 3.6.

Table 3.6. Earnings function for household businesses

	OLS estimation without correcting for sample selection (1)	**OLS estimation with selection correction terms (2)**
VARIABLES	**Household Hourly Business Income (logged)**	
Female Decision Maker	-0.176	-0.236
	(0.196)	(0.198)
Hindu OBC	**-0.160*****	**-0.180*****
	(0.061)	**(0.065)**
SCST	**-0.266*****	-0.180
	(0.076)	(0.173)
Muslim Upper Caste	-0.123	-0.152
	(0.083)	(0.105)
Muslim OBC	**-0.148***	-0.198
	(0.076)	(0.135)
Other	0.069	0.057
	(0.154)	(0.159)
Male head age	-0.022	-0.025
	(0.039)	(0.053)
Male head age (squared)	0.000	0.000
	(0.000)	(0.001)
Female head age	-0.010	0.002
	(0.037)	(0.038)
Female head age (squared)	0.000	0.000
	(0.000)	(0.000)
Highest male education (in years)	**0.031*****	**0.029*****
	(0.006)	**(0.007)**
Highest female education (in years)	**0.020*****	**0.017****
	(0.005)	**(0.008)**
Average secondary class score of the household	0.030	0.029
	(0.069)	(0.070)
Credit	0.077	0.072
	(0.056)	(0.056)
Number of female children (between 5 and 15 y.o.)	-0.017	-0.016
	(0.032)	(0.034)
Number of male children (between 5 and 15 y.o.)	**0.075****	**0.064****
	(0.030)	**(0.031)**
Number of individuals in the household	**0.045*****	**0.040***
	(0.012)	**(0.024)**

Table 3.6 continued on the next page

Table 3.6 (continued)		
Sel_1	N.a	-0.284
		(0.495)
Sel_2	N.a	0.001
		(0.990)
Sel_3	N.a	-0.726
		(0.877)
Manufacturing	**-0.414****	**-0.409****
	(0.202)	**(0.203)**
Services	**-0.603*****	**-0.608*****
	(0.198)	**(0.199)**
Construction	**0.492***	**0.492***
	(0.261)	**(0.261)**
State	Yes	Yes
Constant	**4.400*****	**4.563*****
	(0.511)	**(1.478)**
Observations	1,989	1,977
R^2	0.155	0.159
Robust standard errors (1) and bootstrapped standard errors (2) in parenthesis.		

Source: Author's calculations from IHDS data

The fact that this sector does not contain any specific segments implies that there are no traps inside the household business sector. Male education is a significant determinant of income but the mgnitude of the effect is small. When all other factors are held constant, a one-year increase in male education (for the most educated man in the household) is associated with an increase in hourly business earnings by 2.9%. One way of detecting how women fare in businesses is to consider whether their hourly earnings are higher or lower when they are the primary decision-makers in businesses. We find no significant differences between households in which women are decision-makers and other households. The low presence of female decision-makers in the study sample (1.57%) should also be noted.

Concerning socio-religious groups, when all other factors are held constant, the hourly earnings of Hindu OBCs are 18% lower than Hindu Upper Castes'. All of the other variables concerning socio-religious groups are not significant. Deshpande and Sharma (2016) find significant differences between the business earnings of SCST households and other groups. Nevertheless, they do not control for sample selection. The first column of Table 3.6 shows the same trend. When sample selection is not accounted for, being an SCST or and Muslim OBC significantly and negatively affects income. However, these effects are not robust to the correction of the sample selection which suggests that the smaller business income of these groups is due to barriers to entry into business work rather than discriminative behaviors by clients for instance. Nevertheless, Hindu

OBCs households have lower returns to their personal characteristics, all other factors being held equal. Multiple reasons can explain why this group might face inequality in the household business sector. The existence of discrimination from suppliers or customers may affect their profits. Another explanation of this effect, as suggested by Iyer, Khanna, and Varshney (2011) is the difference between the social networks of each group. The following Table 3.7 shows average social network scores[79] by religion and caste group among households that engage in business work. The first score indicates the extent of social networks within relatives, community and caste and the second score indicates the extent of social networks outside of the community or caste. The maximum score in both cases is 10. Hindu Upper Castes have the most extended social networks, within and outside of the community. Conversely, Muslim OBCs have the smallest networks. The pairwise comparisons tests of equality of means (Appendix 3.4) show that SCSTs have significantly lower "within" social network scores than all groups except for Hindu OBCs and Muslim OBCs. These differences could explain both the selection into business work as well as the significant negative effect on earnings found for Hindu OBC individuals.

Table 3.7. Social network characteristics

Group	Mean score for Social Network 1	Mean score for Social Network 2
Upper Caste Hindu	2.661	3.554
	(2.582)	(3.088)
SCST	1.635	2.718
	(2.172)	(2.644)
Hindu OBC	1.829	2.630
	(2.463)	(2.831)
Upper caste Muslim	2.062	2.728
	(2.277)	(2.849)
Muslim OBC	1.551	2.383
	(1.946)	(2.456)
Other	2.618	3.658
	(2.604)	(3.074)

Source: Author's calculations from IHDS data

[79] The dataset contains information on whether individuals have acquaintances within and outside of the "relatives/ caste/ community". We construct a score for the within community social network (Social Network 1) and one for the outside of the community social network (Social Network 2) by summing the following types of acquaintances: doctors, health workers, teachers, other school workers, public officers, other government employees, elected politicians, police inspectors, other police officers, military officers.

5.2. A segmented salaried employment sector: individual level estimations

5.2.1. Allocation of workers between salaried work and business work

Table 3.8 shows the multinomial logistic estimation of workers' allocation across salaried employment, business employment and no employment (reference group). The coefficients shown in the table are the relative risk ratios (exponentiated coefficients). Two additional estimations are presented with samples of men and women respectively. This estimation is used to compute the selection terms that will be used in the individual earnings functions.

Table 3.8. Multinomial logit estimation of the allocation of workers into occupations

	Whole sample		**Female sample**		**Male sample**	
	Salaried Work	**Business work**	**Salaried Work**	**Business work**	**Salaried Work**	**Business work**
Female	**0.032*****	**0.027*****	-	-	-	-
	(0.001)	**(0.001)**	-	-	-	-
Hindu OBC	**1.247*****	**1.171*****	**1.252*****	**1.405*****	1.089	0.994
	(0.051)	**(0.059)**	**(0.071)**	**(0.126)**	(0.074)	(0.075)
SCST	**1.612*****	**0.776*****	**1.695*****	0.928	**1.307*****	**0.619*****
	(0.070)	**(0.045)**	**(0.100)**	(0.096)	**(0.097)**	**(0.053)**
Muslim Upper Caste	**0.882***	0.938	**0.730*****	1.026	0.929	0.954
	(0.057)	(0.074)	**(0.073)**	(0.150)	(0.097)	(0.111)
Muslim OBC	**0.730*****	0.972	**0.567*****	0.895	0.866	1.127
	(0.044)	(0.070)	**(0.051)**	(0.121)	(0.084)	(0.121)
Other	1.118	**1.441*****	**1.419*****	**1.509****	**0.787***	1.094
	(0.093)	**(0.144)**	**(0.152)**	**(0.270)**	**(0.104)**	(0.158)
Age	**1.321*****	**1.373*****	**1.248*****	**1.196*****	**1.362*****	**1.443*****
	(0.011)	**(0.015)**	**(0.014)**	**(0.023)**	**(0.018)**	**(0.022)**
Age (squared)	**0.996*****	**0.996*****	**0.997*****	**0.998*****	**0.996*****	**0.995*****
	(0.000)	**(0.000)**	**(0.000)**	**(0.000)**	**(0.000)**	**(0.000)**
Primary schooling	**0.718*****	0.921	**0.508*****	0.955	**1.377*****	**1.565*****
	(0.042)	(0.071)	**(0.041)**	(0.113)	**(0.160)**	**(0.203)**
Middle schooling	**0.650*****	0.960	**0.427*****	0.894	**1.154***	**1.575*****
	(0.027)	(0.052)	**(0.025)**	(0.080)	**(0.089)**	**(0.137)**

Table 3.8 continued on the next page

Table 3.8 (continued)						
Secondary schooling	**0.429*** **	0.822	**0.329*** **	0.623	**0.673* **	1.295
	(0.057)	(0.134)	**(0.064)**	(0.209)	**(0.141)**	(0.301)
Tertiary education	**0.552*** **	0.872	**0.633** **	0.603	**0.545*** **	0.947
	(0.075)	(0.147)	**(0.127)**	(0.210)	**(0.117)**	(0.225)
Sec1	**1.581*** **	0.980	**1.694*** **	1.176	1.118	0.693
	(0.217)	(0.166)	**(0.340)**	(0.416)	(0.237)	(0.163)
Sec2	1.179	0.975	1.017	1.185	1.218	0.936
	(0.156)	(0.158)	(0.200)	(0.401)	(0.251)	(0.213)
Sec3	1.098	0.997	1.038	1.473	1.085	0.897
	(0.154)	(0.171)	(0.218)	(0.523)	(0.233)	(0.214)
Married	**0.671*** **	**0.901** **	**0.363*** **	**0.561*** **	**3.363*** **	**4.084*** **
	(0.024)	**(0.045)**	**(0.017)**	**(0.044)**	**(0.240)**	**(0.334)**
NCHILDF	**1.087*** **	**1.170*** **	**0.923* **	1.105	**1.135** **	**1.206*** **
	(0.033)	**(0.043)**	**(0.042)**	(0.077)	**(0.061)**	**(0.070)**
NCHILDM	**1.159*** **	**1.216*** **	**1.084* **	1.032	**1.107* **	**1.192*** **
	(0.035)	**(0.045)**	**(0.047)**	(0.075)	**(0.059)**	**(0.069)**
Number of female children (between 5 and 15 y.o.)	**1.233*** **	**1.227*** **	**1.264*** **	**1.246*** **	**1.268*** **	**1.268*** **
	(0.027)	**(0.034)**	**(0.037)**	**(0.060)**	**(0.056)**	**(0.061)**
Number of male children (between 5 and 15 y.o.)	**1.146*** **	**1.156*** **	**1.125*** **	**1.189*** **	**1.190*** **	**1.189*** **
	(0.026)	**(0.033)**	**(0.035)**	**(0.059)**	**(0.051)**	**(0.055)**
Number of female infants (<5 y.o.)	**0.751*** **	**0.865*** **	1.041	1.064	0.700***	0.831***
	(0.027)	**(0.038)**	(0.052)	(0.088)	(0.037)	(0.050)
Number of male infants (<5 y.o.)	**1.132*** **	**1.266*** **	**1.135*** **	**1.137* **	**0.853*** **	**0.973**
	(0.038)	**(0.054)**	**(0.051)**	**(0.085)**	**(0.047)**	**(0.060)**
Number of elderly male household members	**0.883*** **	**0.894*** **	**0.873*** **	**0.861*** **	**0.883*** **	**0.897*** **
	(0.008)	**(0.010)**	**(0.011)**	**(0.018)**	**(0.013)**	**(0.014)**
Number of elderly female household members	**0.671*** **	**0.901** **	**0.363*** **	**0.561*** **	**3.363*** **	**4.084*** **
	(0.024)	**(0.045)**	**(0.017)**	**(0.044)**	**(0.240)**	**(0.334)**
Number of household members	**1.087*** **	**1.170*** **	**0.923* **	**1.105**	**1.135** **	**1.206*** **
	(0.033)	**(0.043)**	**(0.042)**	**(0.077)**	**(0.061)**	**(0.070)**
Constant	**0.155*** **	**0.009*** **	**0.030*** **	**0.003*** **	**0.057*** **	**0.003*** **
	(0.030)	**(0.002)**	**(0.008)**	**(0.002)**	**(0.017)**	**(0.001)**
Observations	40,525	40,525	20,540	20,540	19,985	19,985
Log likelihood	-29473.697		- 12585.699		-15784.9	
	LR chi2(102) = 22615.88		LR chi2(104) = 2830.16		LR chi2(104) = 4291.06	
	Prob > chi2 =0.0000		Prob > chi2=0.0000		Prob > chi2=0.0000	

Source: Author's calculations from IHDS data.
Note: Exponentiated coefficients. Standard error in parenthesis *** p<0.01, ** p<0.05, * p<0.1

The results show that when all of the other factors are held constant, being a woman decreases the chance of working compared to being inactive or unemployed by about 97%. This result is explained by the large number of unemployed women in the sample.

Compared to Hindu Upper Castes, Hindu OBCs and SCSTs have a higher probability of engaging in salaried work than being unemployed of 24% and 61% respectively. Hindu OBCs also have a higher probability of doing business work than being unemployed or inactive whereas the relative risk ratio for SCSTs indicates the contrary. Muslims have negative ratios which indicate that they are less likely to be in salaried employment than to be unemployed when all other factors are held constant. Indeed, Muslim OBCs are 27% less likely to participate in the salaried sector than to be unemployed in the Whole sample. Yet, the coefficient is only significant in the female sample.

These results indicate important selection into sectors, especially for women. Moreover, since many women are in the reference category even when they are educated. The relative risk ratios of education indicate that compared to individuals who do not have an education, having a higher level of education decreases the odds of being in salaried work. In the male sample, primary and middle schooling are positively associated with salaried work and business work, but secondary schooling and tertiary education are negatively associated with labor market participation. This result supports the "necessity" hypothesis of salaried work, especially among women.

5.2.2. Partition and segment membership

We follow the same steps as for the household business sector to explore the heterogeneity of the salaried sector of the labor market. We estimate a single-segment OLS model and FMMs with different partitions. Note that we add the selection terms estimated at the individual level to control for the sample selection bias.

Table 3.9. Bayesian Information Criterion of the salaried employment models

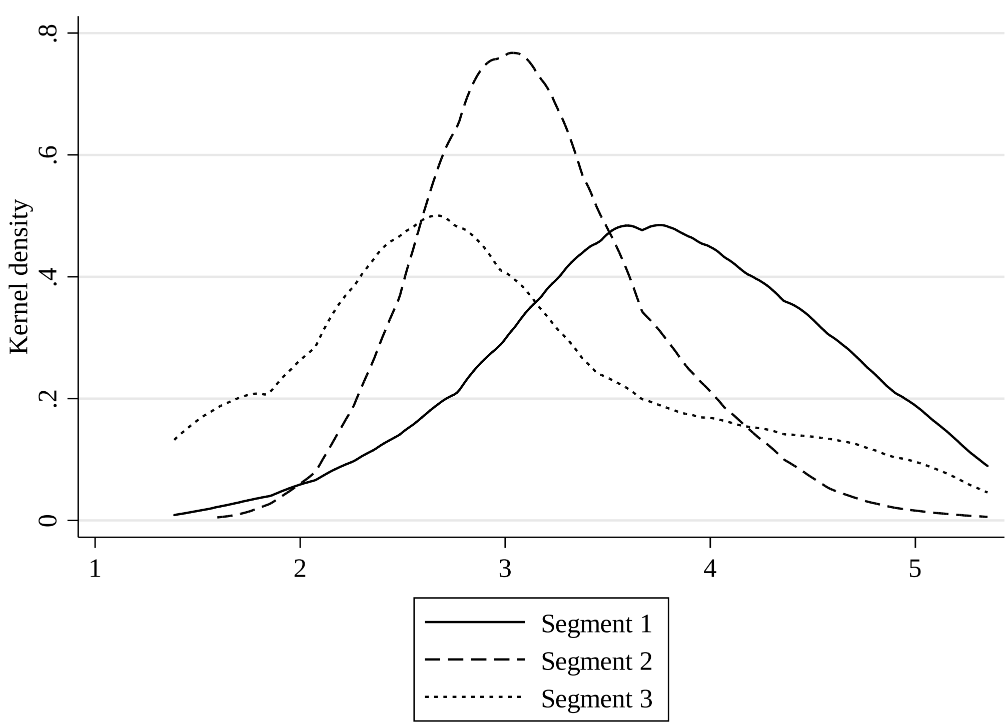
Model	Observations	Log-likelihood	Degrees of Freedom	Bayesian Information Criteria
1-segment (OLS)	14175	-13721.39	55	27968.53
2-segment	14175	-12989.4	119	27116.35
3-segment	14175	-12664.92	180	27050.5
4-segment	14175	-12552.58	223	27236.87
More than 4 segments	*The model does not converge within 150 iterations.*			

Source: Authors calculations from IHDS data

Contrary to the business sector, salaried employment is best described by a segmented approach than by a homogenous one. Indeed, using the Bayesian Information Criterion as an indicator of model quality, we retain the 3-segment model for the rest of the analysis. The smaller BIC for the 3-segment model shows that a model with a mixture of densities is a better fit for the data than a single-segment model. These three segments are determined by the structure of the hourly earnings density function and by the latent class conditioning variables which control for a non-random allocation of workers across the segments based on gender, caste and religion.

The fact that the salaried employment sector is not homogenous implies that there are three types of earnings functions in the labor market, each having its own density function. The kernel density plots in Figure 3.2 show that the density functions of the three segments are indeed considerably different. Segment 1, is the higher wage segment, while Segment 2 is the medium wage segment and Segment 3 is the lower wage segment.

Figure 3.2. Kernel density estimation of segment-specific earnings distributions



Source: Authors calculations from IHDS data

The odds ratios table in Appendix 3.4 show that gender is a strong determinant of the way workers are allocated into the different segments. Women are more likely to be in Segment 3 than in the other two segments. Indeed, the average probability for a woman to be in the third segment is 99%. Interestingly, socio-religious groups only influence the probability of being in Segment 2 compared to Segment 1. Compared to Upper Caste Hindus, all of the groups are more likely to be in Segment 2 than in Segment 1. These results suggest the existence of gender segregation in the labor market, as women are either directed to this specific segment because of employment discrimination or because they choose specific forms of occupations where men are not present. They also show that Hindus are concentrated in the segments with higher earnings.

5.2.3. Earnings structure

Table 3.10 shows the earnings function for the whole salaried employment sector (OLS estimation) and each of its segments from the FMM estimation.

Table 3.10. Earnings functions for the salaried employment sector (OLS and FMM estimations)

Variable	**OLS**		**FMM**	
		Segment 1	**Segment 2**	**Segment 3**
Share of the sector	**100%**	**40.70%**	**37.38%**	**21.93%**
Female	**-0.163***	N.a.	**1.897*****	**0.901*****
	(0.097)	N.a.	**(0.203)**	**(0.321)**
Hindu OBC	**-0.093*****	0.007	**-0.099*****	-0.072
	(0.020)	(0.033)	**(0.032)**	(0.048)
SCST	**-0.097*****	0.039	**-0.119***	-0.089
	(0.030)	(0.071)	**(0.065)**	(0.060)
Muslim Upper Caste	**-0.067****	-0.049	0.004	**-0.261*****
	(0.032)	(0.057)	(0.045)	**(0.090)**
Muslim OBC	-0.031	0.037	-0.049	-0.013
	(0.030)	(0.085)	(0.048)	(0.078)
Other	0.043	0.095	-0.093	0.072
	(0.040)	(0.059)	(0.107)	(0.076)
Age	**0.032*****	**0.041****	0.011	**0.051*****
	(0.008)	**(0.017)**	(0.015)	**(0.018)**
Age squared	**-0.000****	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Primary education	**0.140*****	**0.137*****	0.041	**0.140****
	(0.023)	**(0.048)**	(0.036)	**(0.060)**
Lower secondary education	**0.206*****	**0.193*****	0.042	**0.280*****
	(0.022)	**(0.047)**	(0.042)	**(0.050)**

Table 3.10 continued on the next page

Table 3.10 (continued)				
Upper secondary education	**0.328*** **	**0.375*** **	**0.097* **	**0.387*** **
	(0.028)	**(0.052)**	**(0.052)**	**(0.072)**
Tertiary education	**0.704*** **	**0.856*** **	**0.182*** **	**0.942*** **
	(0.024)	**(0.041)**	**(0.043)**	**(0.065)**
SSC First Class	**0.334*** **	**0.335*** **	**0.116** **	**0.280*** **
	(0.022)	**(0.026)**	**(0.056)**	**(0.051)**
Married	**0.108*** **	**0.109*** **	**0.131*** **	0.039
	(0.019)	**(0.036)**	**(0.033)**	(0.039)
Number of daughters (<15 y.o.)	**-0.025*** **	0.004	-0.023	-0.022
	(0.008)	(0.012)	(0.017)	(0.013)
Number of sons (<15 y.o)	**-0.014* **	-0.001	-0.010	0.000
	(0.008)	(0.011)	(0.017)	(0.000)
Manufacturing	**0.106*** **	**0.203*** **	0.058	-0.056
	(0.026)	**(0.070)**	(0.055)	(0.047)
Services	**0.161*** **	**0.343*** **	**-0.097* **	**0.103*** **
	(0.025)	**(0.069)**	**(0.056)**	**(0.040)**
Public Administration	**0.504*** **	**0.338*** **	**1.054*** **	**0.520*** **
	(0.034)	**(0.075)**	**(0.128)**	**(0.083)**
Construction	**0.222*** **	**0.253*** **	**0.143*** **	**0.263*** **
	(0.026)	**(0.069)**	**(0.051)**	**(0.043)**
Sel1	**0.641*** **	-0.273	**0.987** **	**0.698** **
	(0.158)	(0.535)	**(0.411)**	**(0.281)**
Sel2	0.071	-0.508	0.314	0.044
	(0.057)	(0.403)	(0.258)	(0.121)
Sel3	**0.529*** **	-0.862	0.731	**1.294* **
	(0.197)	(0.678)	(0.527)	**(0.700)**
State	Yes	Yes	Yes	Yes
Constant	**2.345*** **	**1.720*** **	**3.113*** **	**1.025** **
	(0.154)	**(0.336)**	**(0.290)**	**(0.406)**
Segment Variance		0.279	0.109	0.456
		(0.017)	(0.012)	(0.018)
R^2	0.403		N.a.	

Source: Authors calculations from IHDS data.
Note: Household-level clustered standard error in parenthesis *** p<0.01, ** p<0.05, * p<0.1

The OLS estimation for the whole labor market shows that overall the returns to educations are linear. Compared to workers who have no education, there is an education premium of 14%, 20.6%, 32.8% and 70.4% or primary, lower secondary, upper secondary and tertiary education respectively. For an equivalent level of education, returns to ability are also significant because individuals who have a first-class score in the SSC benefit from 33.4% higher hourly wages.

As shown in Table 3.11, when workers are allocated in segments where they are the more likely to be in, women are almost exclusively in Segment 3 and men are almost exclusively in Segments 1 and 2.[80] We can, therefore, consider that there is a "*Female Segment*" (Segment 3) an "*Upper Male Segment*" (Segment 1) and a "*Lower Male Segment*" (Segment 2).

The OLS estimation shows that when all other factors are held constant, being a woman significantly impacts hourly wages by -16.3%. Since men and women are in different segments of the labor market, this negative effect can be explained by the fact that both groups do not hold the same types of occupations (even when they work in the same industries). Therefore, their earnings structure is different and the average earnings in the female segment (Segment 3) is smaller than the average of the two other segments combined. In the Female Segment, the returns to education are higher than in the other two segments, and they increase as the level of education increases. It is important to note that the fact that the Female Segment has higher returns to education does not imply that women earn more than men when they are more educated. Indeed, since the earnings density function of women is more skewed to the right than the ones for men, this difference means that relative to the earnings in the female segment, women who are more educated benefit from higher premiums. The negative effect of being from the Hindu OBC or SCST group (average -9.3% and -3.7% compared to Hindu Upper Castes respectively) in the OLS estimation is not present in every segment. In this Upper Male Segment, the differences between group earnings are mostly due to education differentials. In the Lower Male Segment (Segment 2), belonging to the Hindu OBC or SCST group significantly and negatively affects hourly wages by about 10%. The significance level for the SCST dummy variable is lower than in the OLS estimation (10% instead of a 1% level). In other words, Hindu Upper Caste individuals are concentrated in the Upper Male Segment, but those who are in the Lower Male Segment also earn significantly more than SCSTs or Hindu OBCs.

From the FMM estimation, it is possible to compute the probability for each worker to be part of each segment. In order to further analyze their characteristics, we allocate each worker to the segment for which their *membership probability is the highest*. Table 3.11 provides statistics that will allow describing the segments.

---

[80] Consequently, the coefficients for the dummy variable "*Female*" are irrelevant since there are either no or very few women in Segments 1 and 2 or there are very few men in Segment 3

The segments do not reflect clear boundaries regarding the type of employment. Formal employment, in the Indian legal framework, is almost exclusively limited to permanent workers (Fagernäs 2010), who are in the group "*Long-term contract*" of the table. However, this group does not only contain formal workers as some of them may not have any form of employment contract. Although individuals who hold this type of contract are in the most remunerating segment (50.87%), they are also present in the two other segments, suggesting that all long-term contract holders do not necessarily have the same earnings function. Conversely, casual labor, although more present in the Segment 2 is also divided among the three groups. Segment 2 is more precarious with more frequent daily payments. On the contrary, the first segment has workers with the most frequent monthly payments. In terms of social network, there is a clear difference between the segments. The lower male segment has lower scores for the intra-community social network as well as the inter-community one. Social network can help individuals access better occupations. In India, Gille (2018) shows that when members of specific *jatis* are elected as officials, there is an increase in affirmative action policy applications, probably shaped by beliefs. Networks can also be an outcome of the segment membership, and a larger social network (Segment 1) may be beneficial in the long-term.

Table 3.11. Segment membership

	Segment 1 "Upper Male Segment"	Segment 2 "Lower Male Segment"	Segment 3 "Female Segment"
	Percent of the labor market (Row total = 100%)		
Hourly wage in INR	57.69	28.08	32.67
(Std. Dev.)	(47.76)	(20.58)	(40.06)
Women (%)	0	0.49	99.51
Hindu Upper Caste	63.47	14.76	21.77
Hindu OBC	32.35	45.61	22.05
SCST	33.52	41.68	23.79
Muslim Upper Caste	40.90	44.04	15.06
Muslim OBC	16.50	67.91	15.59
Other	59.88	11.41	28.72
Industry			
Agriculture	26.75	33.33	39.92
Manufacturing	38.81	48.47	15.71
Services	42.49	32.01	25.50
Public administration	62.05	25.23	12.72
Construction	29.47	56.04	14.49
Type of employment			
Casual daily	31.11	48.02	20.87
Casual Piecework	29.42	43.01	27.57
Regular employment (< 1 year)	42.43	34.73	22.83
Long term employment (>1 year)	50.87	27.34	21.79
Pay-rate			
Per day	30.01	51.39	18.60
Per month	47.57	30.34	22.09
Social network score 1	2.33 (2.56)	1.56 (2.10)	1.96 (2.49)
Social network score 2	3.29 (3.04)	2.39 (2.59)	2.79 (2.91)

Source: Authors calculations from IHDS data

5.2.4. Robustness verifications

A first question that arises concerning the segmentation analysis for the salaried sector is whether education and productivity variables are significantly different across the segments or not. Indeed, finding significant differences in the coefficients that measure the returns to education is a confirmation that the groups we detect are relevant segments. Tests of equality of coefficients (i.e. contrast tests) show significant differences for all variables (Table 3.12).

Table 3.12. Equality of coefficients tests

Variable	Test	Chi-square
Primary education	Segment 1 versus Segment 2	**4.45****
	Segment 2 versus Segment 3	**2.94***
	Joint	**5.46***
Lower Secondary School	Segment 1 versus Segment 2	**11.66*****
	Segment 2 versus Segment 3	**17.48*****
	Joint	**20.04*****
Upper Secondary School	Segment 1 versus Segment 2	**19.82*****
	Segment 2 versus Segment 3	**11.87*****
	Joint	**21.92*****
Higher education	Segment 1 versus Segment 2	**124.52*****
	Segment 2 versus Segment 3	**90.41*****
	Joint	**146.95*****
SSC Class 1	Segment 1 versus Segment 2	**8.98*****
	Segment 2 versus Segment 3	**4.51****
	Joint	**9.01****

Source: Authors calculations from IHDS data

Furthermore, the FMM results we presented take into consideration the fact that gender, religion and caste condition the allocation of workers into different segments. The relevance of this assumption can be discussed. The following arguments justify our choice of adding conditioning variables. Not adding predictor variables supposes that the segmentation is exogenous. However, considering this exogeneity when there are strong links between social identity and the allocation of individuals across occupations is likely to yields biased results. The comparison of the BIC of the 3-segment model with and without conditioning variables indicate that the model with the conditioning variables is a better fit for the data.

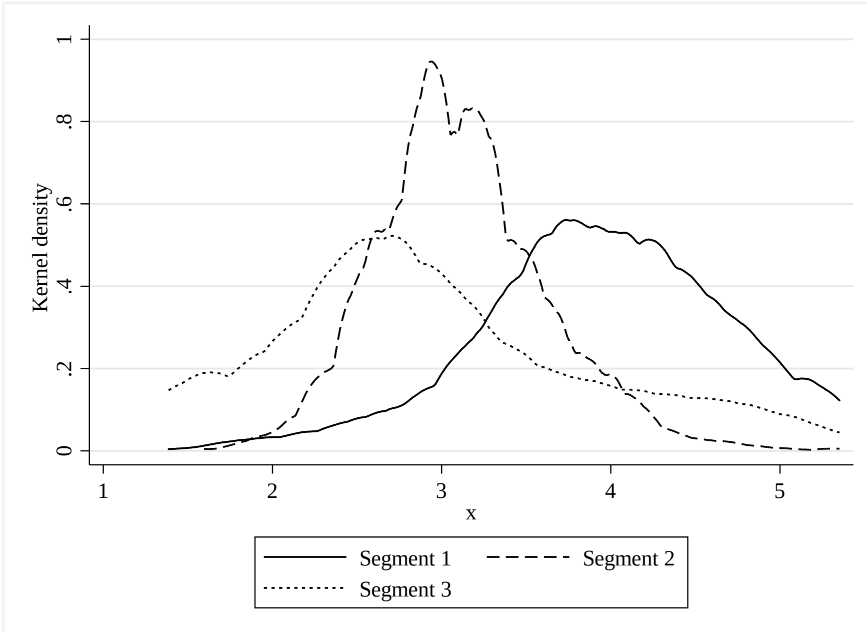
Table 3.13. Bayesian Information Criterion comparison

Model	Observations	Log-likelihood	Degrees of Freedom	Bayesian Information Criterion
3-segment with conditioning variables	14175	-12664.92	180	27050.50
3-segment without conditioning variables	14175	-12716.38	170	27057.83

Source: Authors calculations from IHDS data

Furthermore, Table 3.14 and Figure 3.3 shows that an estimation in which we add the SSLC first-class variable as an additional predictor variable, shows similar results in terms segment allocation. Almost all women are still allocated into one single segment.

Figure 3.3. Segments from alternative specification



Source: Author's calculations from IHDS

Table 3.14. Gender distribution across segments in alternative specification

	Segment 1	Segment 2	Segment 3
	Percent of the labor market (Row total = 100%)		
Women (%)	0	1%	99%
Men (%)	42%	53%53	5%

Source: Author's calculations from IHDS

5.3. Structural traps, Formal *versus* Informal and Necessity *versus* Opportunity: insights from our estimations

The results we have presented in sections 5.1 and 5.2 point the segmented nature of the urban Indian labor market in the salaried employment sector and as a whole. In a context of predominant informality of production units and employment, these results can shed light on the relevance of a Formal *versus* Informal debate and a Necessity *versus* Opportunity debate in the caste of India.

The household business sector is exclusively composed of small-scale production units (between one and seven workers) which do not require to be registered and fall into the Indian definition of unregistered enterprises. Adapting the formal/informal issue to Indian context would imply that household businesses have different patterns of earnings structure and that some of them would be formal micro-enterprises, had the threshold for regulation been lower than 10 workers. However, the fact that we find a homogeneous sector does not support this hypothesis. Heterogeneity in this sector is more likely to take a continuous and linear nature (e.g. a linear relationship between the level of capital invested and income) than a dichotomous one. Furthermore, the opportunity/necessity dichotomy opposes production units that choose to stay informal to avoid registration costs to those who have to operate informally by necessity. In the case of the household business sector, the fact that none of the businesses in the sample have more than seven employees, even when they are not household members suggest that there are no or very scarce businesses that operate informally for opportunity reasons.

In the salaried sector, which is also predominantly informal, one way to adapt the question of the formal/informal divide is to observe whether the allocation of individuals in segments reflects a deliberate choice. In this case determining whether workers are in the segment where they maximize their earnings or not can reveal the existence of traps[81]. Concerning salaried work, the estimations show the existence of three segments. Günther and Launov (2012) and Salem and Bensidoun (2012) test the relevance of the segments they find with the FMM with a similar test. To compute the predicted earnings, we exclude gender and socio-religious variables. The results presented in Table 3.16 and Figure 3.4. show that in all cases workers would be maximizing their earnings in the segment 1. With their personal characteristics, women would benefit from higher earnings if they were in segment 1 or 2.

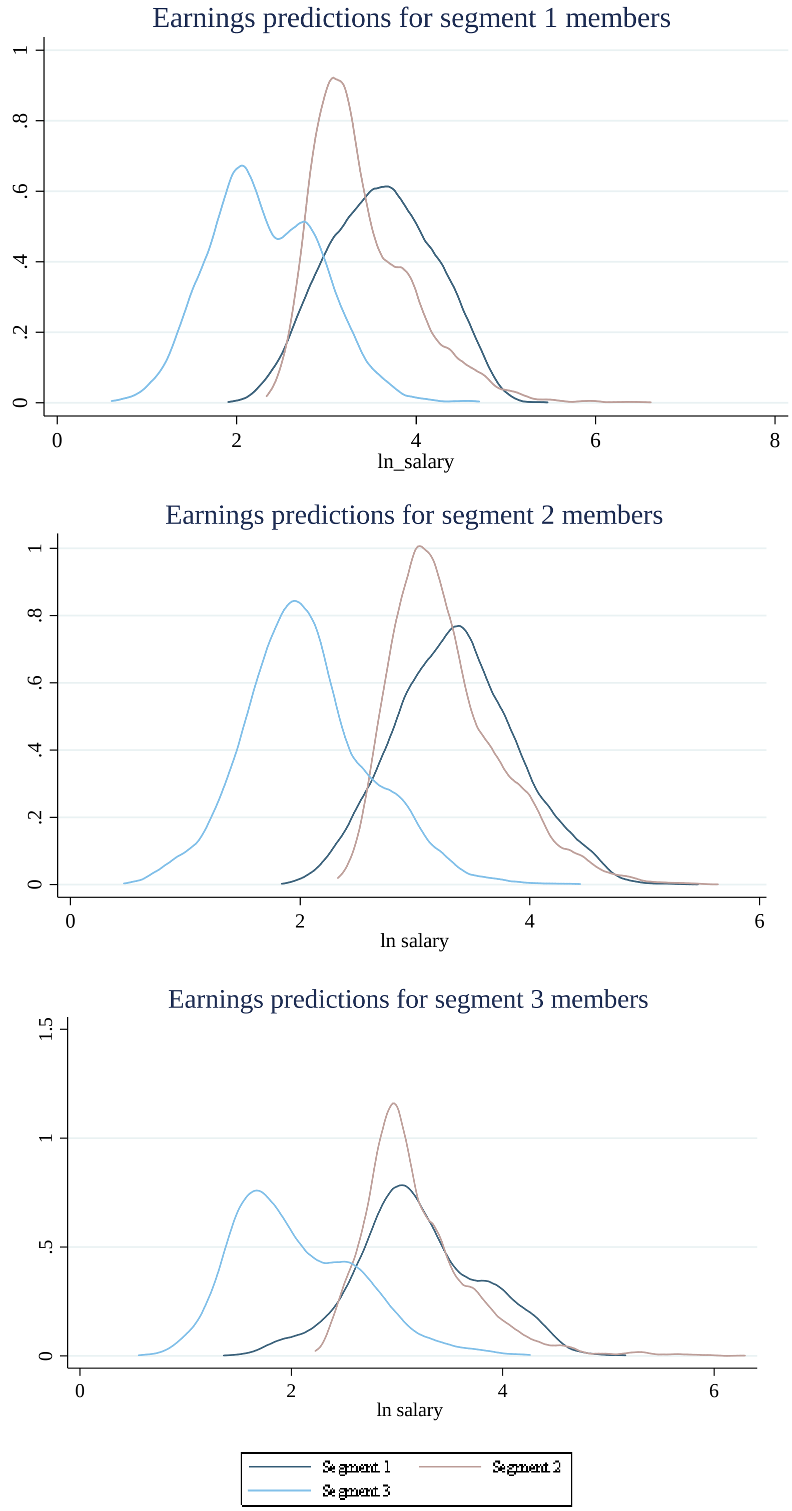
Table 3.15. Predicted average earnings by segment (average)

	Segment membership		
	Segment 1	**Segment 2**	**Segment 3**
Predicted average log earnings in Segment 1	**3.610 (0.593)**	**3.357 (0.531)**	**3.205 (0.589)**
Predicted average log earnings in Segment 2	3.396 (0.568)	3.280 (0.479)	3.176 (0.508)
Predicted average log earnings in Segment 3	2.318 (0.610)	2.053 (0.546)	2.048 (0.593)

Source: Authors calculations from IHDS data

[81] Note that the hypothesis that workers wish to maximize their earnings is made in this demonstration. Other rationales may be as relevant, such as wishing to maximize employment stability or quality. These analyses would require a separate study.

Figure 3.4. Predicted earnings by segment (distribution)



Source: Authors calculations from IHDS data. *Note*: *ln_salary* in the x-axis refers to logged hourly earnings

These results suggest that Segment 2 and 3 are structural traps. In a homogenous labor market, there would be only one segment with similar returns to personal characteristics for all workers, which is not the case in the salaried sector in India. The Upper Male segment (Segment 1) contains the higher shares of potentially formal employment with more regular and permanent jobs, monthly pay rates and a higher wage. The Lower Male segment which contains higher shares of potentially informal employment: casual labor, daily pay rates and lower wages. The opposition of both segments suggest the existence of a stratification in the Indian labor market. However, the distribution of casual, regular, and permanent employment across the segments also suggest that the formal/informal divide is best represented by the idea of a continuum.

The third form of segmentation that exists in the urban Indian labor market is between salaried work and self-employment. Although, we only address the issue as a selection bias problematic in this study, this form of segmentation can also shed light on the formal/informal divide. Indeed, a particularity of the Indian labor market is that informality can take very heterogeneous forms. Although self-employment is often the type of informality depicted as the economy of the poor, the existence of large shares of contract labor constitutes another form of very precarious work.

6. Conclusion

By considering segment membership as a latent variable, we estimate the number of segments with a finite mixture model aimed at identifying different distributions of income functions. The labor market segmentation that we uncover reveals patterns of informality and allows us to better understand the phenomenon than with a possibly incorrect binary distinction of formal and informal activities. Focusing on segmentation yields interesting results concerning the mechanisms of inequality in the relationship between educational attainment and labor market outcomes. A novelty of our methodology is to consider the allocation of workers in given segments as non-random. We stipulate that this allocation is determined by gender-, religion- and caste-related cultural norms. The estimation results allow us to understand the inter- and intra-segment role of these sociodemographic characteristics. We conclude by discussing the relevance of *formal* versus *informal* an *opportunity versus necessity* debate with respect to India’s informal sector.

We find that the household business sector is homogeneous. Barriers of access in this segment particularly affects SCST households whose income are consequently smaller. Horizontal inequalities however persist in this segment since Hindu OBCs have lower incomes, even after controlling for educational and personal characteristics. This estimation does not allow to analyze the gender differentials because of the scarcity of female decision-makers in the sample. Being a decision-maker in the household is far from being the only way women participate in household businesses. They are mostly contributing family members. There is a probable underestimation of this form of employment which is particularly vulnerable (UNDP 2015; ILO 2016b). Indeed, only 13.34% of women are considered to be workers in the business. Moreover, this underestimation can affect the dependent variable of the household earnings estimation, since the denominator is the hours worked in the household.

We also find that the salaried sector is divided into three segments. All women in the labor market are pooled in one segment and men are separated between a lower segment and an upper one. The method we use in this study is a data-driven one, which has the benefit of letting the data speak for itself. We did not, for instance, expect the model to place all women in the same segment. An important contribution of this study is therefore to offer an empirical grounding to the separation of the male and female earnings structure in analyses concerning urban India. However, the fact that the method is data-driven also has its limitations. Although we do have to determine *a priori* the criteria of a segment in the labor market, the choice of variables that constitute the earnings function remains an arbitrary choice.

In a nutshell, this study confirms that gender-based labor market segregation is a structuring factor of the urban Indian economy. According to Chen (2006), gender segregation is a characteristic of the informal sector, where women have different activities and different earning levels than men. In the predominantly informal labor market of India, it consequently comes to no surprise to find the existence of gender segregation.

Chapter 4. Insights on potential discrimination from the decompositions of wage gaps

1. Introduction

Labor-market discrimination is one of the most universally and firmly condemned forms of work-related stigmas. Nonetheless, it remains hard to prove (Tomei 2003). The ILO defines discrimination as "*any distinction, exclusion or preference made on the basis of race, colour, sex, religion, political opinion, national extraction or social origin, which has the effect of nullifying or impairing equality of opportunity or treatment in employment or occupation*" (Discrimination (Employment and Occupation) convention, ILO 1958). In a labor market perspective, this definition implies that discrimination comes from the demand side. It involves the action of a person or an institution which discriminates (e.g. the employer) and whose behavior creates an opportunity inequality or affects a labor market outcome. Two principal types of discriminatory behavior are emphasized in the labor market literature: employment discrimination (i.e. not employing a person based on a personal characteristic) or wage[82] discrimination (i.e. offering a different remuneration to a worker because of this characteristic). Labor market discrimination can be directed towards a group on the basis of different individual characteristics that can be innate (e.g. gender) or acquired (e.g. political opinion).

In the case of India, discrimination on the grounds of gender, religion and caste call for particular attention given the extent to which these characteristics shape identities. Several issues, such as selective abortion and the preference for sons in household expenditures, point towards a generalized form of discrimination against women in India. The contrast between India's economic development and structural change on the one hand, and the low level of FLMP on the other hand, raises the question of the place of women in the labor market in terms of gender wage parity.

[82] Note that in this chapter, which exclusively analyzes the situation of salaried worker, the terms wages and earnings refer to income from a salaried occupation.

Moreover, violence against specific groups, which can be considered as the most extreme form of discrimination (Sharma 2015), has been regularly documented. Muslim-Hindu political tensions have resulted into regular communal riots and altercations leading up to approximately 40,000 casualties since the partition with Pakistan (Santosh 2015). Violence against SCSTs is also a common occurrence in India. The qualification of *Dalits* as "untouchable" because of the polluting nature of contact with them has motivated centuries of discrimination taking the form of spatial segregation and strong economic deprivations. A report of the National Commission of Scheduled Castes and Scheduled Tribes from 1997 (cited by Sharma (2015)) points out that movements of *Dalit* emancipation were regularly countered by episodes of violence taking the form of mass killings, gang rapes and looting of *Dalit* villages. A growing literature shows the linkages between conflict and economic deprivations in India. For instance, Mitra and Ray (2014) show that the per capita increase of Muslim expenditures is associated with episodes of violence and Sharma (2015) shows that inter-caste equality can have perverse effects because when inequality increases between SCSTs and Upper Castes, violence against SCSTs diminishes. Putting these findings into perspective, similar motivations of maintaining the hierarchy of socio-religious groups may also lead to labor market discrimination.

The existence of stigma against specific groups in India is acknowledged by public institutions through two types of policies. First, quota-based affirmative action policies in favor of SCSTs ensure their access to public employment, higher education and political seats in Union, State and local administrations. Reservations in public employment also exist for Hindu and Muslim OBCs. Second, the Constitution prohibits discrimination on the grounds of religion and caste. Despite these policies, horizontal inequality persists and the Indian society remain substantially stratified in terms of wealth, Hindu Upper Castes being well above Hindu OBCs and non-Hindus, followed by SCSTs (Zacharias and Vakulabharanam 2011).

Since discrimination is difficult to detect in household data, the analysis of wage gaps relies on the assumption that the part of the differential between two groups that is not due to differentials in personal characteristics is *potential discrimination*. One of the main challenges in a wage gap analysis is to account for the possibly non-random sorting of workers into specific labor market outcomes, such as salaried employment or unemployment. A wage differential between two equally productive groups can wrongly be attributed to wage discrimination if individuals have unequal access to occupations. In this case, the wage differential is due to one of the groups having higher-paying jobs than the other. Employment discrimination (whether an employer decides to hire a person or not based on a personal characteristic) or personal choice (that can also be qualified as *self-discrimination*) can lead to such sorting.

The contribution of this study is to shed new light on the existence, the extent and the nature of wage gaps between religion, caste and gender groups by drawing on multiple methodologies. First, we explore wage gaps at the mean using parametric decomposition methods, namely Oaxaca-Blinder and its extension proposed by Fortin (2008). By comparing the results from estimations with and without a correction for the non-random selection into employment, we can detect whether the composition of wage gaps is linked to the allocation of workers across different employment situations. The sample selection correction in decomposition studies in India is usually done with a Heckman method that uses a correction term (see for instance Agrawal (2014)). We estimate parametric decompositions using the Dubbin and MacFadden's correction method developed by Bourguignon, Fournier, and Gurgand (2007).

We further investigate the extent and source of wage gaps in two additional steps. An issue raised by Ñopo (2008) is that the shares of the explained and unexplained wage gaps can be biased because of a lack of common support (i.e. a lack of comparable individual across groups). For this reason, we propose to explore the relevance of this issue in the case of urban India, using Ñopo's method. We compare findings from the parametric decompositions to those yielded by the non-parametric decompositions which consist in comparing wages between matched individuals. This step allows us to determine whether the presence of individuals who cannot be compared to their counterparts from another socio-religious group in the sample contributes to the wage gap. Furthermore, we proceed to the analysis of wage gaps across the distribution in a descriptive manner by comparing the matched samples generated by the Ñopo decomposition method and then by estimating the explained and unexplained parts of the wage gap using quantile decomposition methods.

2. A literature review on the analysis of wage gaps

Labor market discrimination refers to the situation where two individuals are treated unequally because of an observable characteristic despite being equally productive. The unequal treatment can take many forms but the most commonly analyzed ones, in the labor market literature concerning gender or socio-religious groups, are wage and employment discrimination. Speaking about discrimination requires ruling out labor market outcomes that are due to differences in human capital endowments, productivity, or preferences. The latter being especially hard to observe, studies analyze wage discrimination by introducing the concept of *potential discrimination*, which is the share of the wage gap between two groups that remains after controlling for education, productivity-related characteristics and control variables.

2.1. Wage discrimination from a theoretical perspective

From the perspective of human capital theorists, a wage differential between two individuals reflects their unequal productivity which is caused by a different level of human capital. This framework does not allow to explain a wage differential between two equally productive individuals. To apprehend this type of situation, extensions of the human capital theory consider two types of discrimination: taste-based discrimination and statistical discrimination (Altonji and Blank 1999).

Taste-based discrimination is a form of behavior emanating from the employer, coworkers or consumers (Becker 1971). It may take the form of employment discrimination and/or wage discrimination. In this framework, employer discrimination takes place when an employer discriminates against (or favors) a worker because of his distaste (or taste) of working with someone belonging to a specific demographic group (i.e. based on gender, caste, age, sexual orientation, etc.). Suppose that there are two groups of workers, A (the majority group) and B (the minority – and disadvantaged – group). A discriminative behavior results in not employing the person from group B or employing this person with a lower wage than other workers. Employers (from the A group) perceive the distaste of working with a member of the discriminated group as a supplementary cost, which is why a higher a distaste in working with an individual from group B will be associated to a lower the wage for this group. In this framework, the employer being a profit maximizer, he will stop this behavior in the longer term to stay competitive. Apart from employer discrimination, employee discrimination and consumer discrimination can also lead to wage gaps. An employer can choose not to employ a person because other employees have a distaste in working with them or because consumers have a distaste in purchasing goods or service based on a personal characteristic.

Statistical discrimination refers to a subtler mechanism. In a context of imperfect information on the productivity of workers, an employer will use observable characteristics of a worker (e.g. gender) as a proxy for unobservable but productivity-related characteristics. Models based on the work of Arrow (1973) consider that employers do not discriminate because of preference or distaste. Instead, they estimate the productivity level of their employees by relying on “group statistics” which can be real or based on their perception (Fang and Moro 2011). If an employer thinks that a specific group is less productive (because of a reputation of laziness or because of statistics on the level of education for instance), the employer will choose not employ members of this group while remaining “rational.” Another group of models based on the theory of Phelps (1972) explain the existence of discrimination when there are two equally productive groups. The disadvantaged group may face discrimination because of measurement error affecting this group in particular. Coate and Loury (1993) point out that self-fulfilling stereotypes may lead to initially equally productive groups to become unequally productive.

Apart from what is considered “*pure wage discrimination*,” wage gaps can also be the consequence of an unequal distribution of individuals across different occupations, sectors, of firms. Gender-based occupational segregation can de be caused by differentials in education but also by personal preference (e.g. more flexible hours for child-care) or to conform to a social norm (Ponthieux and Meurs 2015). Furthermore, human capital theories of discrimination underestimate the role of premarket discrimination as they consider education as exogenous (Figart 2009). Bergmann (1974) proposes to analyze the issues of labor market supply and demand simultaneously as these “dynamics of choice and constraint cannot be isolated from each other”. The model shows that discrimination against specific groups (based on gender and race) creates occupational segregation. Workers from these groups are concentrated in specific occupations which therefore become “crowded” and this phenomenon pushes wages down.

2.2. Empirical findings in the Indian context

The interest of labor economists in the gender wage gap has increased in the past decades, especially in emerging countries such as India. Several factors can explain the wage gaps between men and women. Differentials in the level of education are one of the potential causes of wage differentials. Although there is an important gap in literacy rates (Sundaram and Vanneman 2008) and educational attainment, studies show that a part of the gender wage gap remains unexplained (Kingdon and Unni 2001; Menon and Rodgers 2009; Deininger, Jin, and Nagarajan 2013; Deshpande, Goel, and Khanna 2017). Studies at the national level show that the part of the wage gap attributable to potential discrimination has increased over time (Mukherjee and Majumder 2011; Deshpande, Goel, and Khanna 2017). Moreover, using a quantile decomposition framework, Deshpande, Goel, and Khanna (2017) show the existence of a "sticky floor" situation which can be defined as discrimination at the smallest percentiles of the distribution. They also find a less important but existing "glass ceiling" which is defined as discrimination at the higher percentiles of the distribution. Apart from differentials in productive ability or potential discrimination, other factors might cause the gender wage gap as well. In an article in which they revisit the "Boserup paradox,"[83] Mahajan and Ramaswami (2017) point out that female labor supply has sizeable effects on female wages but not on male wage, thus creating a gap between both groups.

Concerning socio-religious groups, Thorat and Attewell (2007) show the existence of caste favoritism and social exclusion of SCSTs and Muslims in the formal private sector. Madheswaran and Attewell (2007) decompose wage gaps between SCSTs and Hindu Upper Caste using nationally representative data. Their results show that about 85% of the wage gap can be explained by human capital endowments, but 15% of the gap remains unexplained and can be attributed to potential discrimination. Approximately one-third of the unexplained portion of the wage gap is due to overpayment of forward castes and two-thirds are due to underpayment of SCSTs. However, they do not control for the sample selection bias. By controlling the sample selection bias, Deininger, Jin, and Nagarajan (2013) find that the hypothesis of no discrimination against SCs cannot be rejected. Using the first wave of the IHDS data, Agrawal (2014) shows that, after

[83] This paradox stems from the work of Boserup (1970) in which she analyzed the gender wage gap in India in the 1950s. She finds that the gender wage gap is higher in the southern States of India, where FLMP is higher.

controlling for the sample selection bias, wage gaps between SCSTs and non-SCST are due to endowment differentials. Ito (2009) proposes a simultaneous estimation of job discrimination and wage discrimination among regular salaried workers in rural North India. He finds no evidence of wage discrimination against SCSTs. The wage gap between SCSTs and the other groups is partly due to their lower educational attainment. Moreover, the group faces higher transaction costs regarding entry into the labor market which is likely to be linked to employment discrimination or smaller or less efficient networks. He concludes that traditional Oaxaca-Blinder decompositions in the case of India tend to overestimate wage discrimination because they do not account for labor market selection. His results concur with the experimental study from Banerjee et al. (2009) which compares callback rates after sending fictitious CVs to job offers in software and call-center jobs in Delhi. Overall, they find no evidence of differential callback rates between Hindu Upper Castes and lower castes (Hindu OBC and SCST) and Muslims. Interestingly, they find important differences in callback rates among male applications for call-center jobs, whereas no differential is found among female applicants. These results suggest that the interaction of religion, caste and gender might sometimes have a compensating effect concerning employment discrimination. Overall, studies which consider the sample selection bias tend to reject the existence of wage discrimination against SCSTs and Muslims.

3. The methodology to analyze wage gaps

The linear relationship between education, productivity and wages which is emphasized in the human capital theory only partially describes the reality of labor markets. In countries such as India, where social norms highly override the rule of law in influencing individual behavior, there is an important leeway for discrimination leading to persistent horizontal inequalities. Labor market discrimination can occur at many levels. It can affect human capital accumulation and lead to productively different groups, it can also affect employment decisions leading to occupational segregation or affect wages. In this section, we will present parametric and nonparametric decomposition methods that will be implemented to estimate the extent and the nature of religion, caste and gender wage gap in urban India.

This section describes the methodology to decompose and analyze wage gaps in order to discuss the hypothesis of potential discrimination between specific groups. The study of income differentials usually requires the use of Mincer-type earnings functions (Eq. 4.1) in which the dependent variable is the natural logarithm of an income variable (y_{ij}). The independent variables (X_{ij}) are individual educational and productivity-related characteristics and control variables.

$$lny_{ij} = \alpha_j + X_{ij}\beta_j + \varepsilon_{ij} \quad \text{[Eq. 4.1]}$$

Observing wage gaps provides interesting observations regarding group differentials in endowments, the main challenge being the choice of a relevant reference earnings function which would prevail in the absence of discrimination (Neumark 1988). Indeed, since there is no true counterfactual (e.g. we cannot truly know what would a man's wage be if he were a woman), we can only posit that one reference earnings function is the one that would prevail in the labor market if there were only one group. For example, we can consider that the male earnings function is the one that would prevail if the labor market contained only men, and compare women earnings function to it, following which we decompose the wage gap. We propose to compare the results from three types of reference earnings functions: the one corresponding to the non-discriminated group, the one corresponding to the discriminated group and pooled earnings function. These decomposition results can, however, be biased because of a lack of common support. In order to see whether this is the case, we compare the results to a non-parametric decomposition method, following which we describe the wage gap across the distribution.

3.1. Dealing with the selection bias

Since the analysis is based on wage-earners, the coefficients of the earnings functions are likely to be biased if individuals face unequal access to salaried occupations. As in chapter 3, we use the method developed by Bourguignon, Fournier, and Gurgand (2007)[84] which estimates the probability for an individual to engage in salaried work instead of being self-employed or inactive. This method yields three selection terms ($\theta selP_{1,}\ \theta selP_{2,}\ \theta selP_3$) that will be added to the equations of interest.

Ben Yahmed (2018) shows that this method can be used to correct the selection bias in decomposition frameworks. She analyzes the unequal sorting of workers into inactivity, the formal

[84] The correction terms estimated in chapter 3 will be used to implement the method.

and informal sectors before applying a wage decomposition method in the context of Brazil. To our knowledge, no decomposition studies concerning the gender, religion and caste gaps in India account for complex selection into the labor market.

3.2. Oaxaca-Blinder decomposition method and its parametric alternatives

While Mincer regressions provide interesting information on the significance of variables such as gender or socio-religious group in the determination of wages, their main drawback is to consider that the wage structure is the same for the comparison groups. This caveat can be overcome by using decomposition methods which allow the intercepts and the coefficients to be different for each group. Moreover, a wage decomposition provides a clear estimation of the “explained” component of the gap, which is the share of the wage gap that can be attributed to differences in observed productive characteristics and the “unexplained” component of the gap which corresponds to potential discrimination (also respectively addressed as the characteristics effect and coefficient effect).

The detection of wage discrimination can be done by a counterfactual methodology allowing a distinction between the role of productive differentials and unexplained differentials from one group to another (Oaxaca 1973, Blinder 1973). This method is usually referred to as the Oaxaca-Blinder (OB) decomposition. In this framework, the unexplained part of the wage gap represents the differential in the way personal characteristics are rewarded on the labor market between two groups. This part of the gap is considered as a potential measurement of discrimination. Other factors than discrimination such as labor-market segregation or unobservable characteristics such as preferences can also be a part of the unexplained part of the wage gap. Nonetheless, despite the uncertainty related to the meaning of the unexplained component (Kingdon and Unni 2001), results from decomposition analyses remain highly informative on the differentiated rewards of human capital between groups.

The basic decomposition framework is the following. α is the intercept and j refers to the group variable. For the sake of the demonstration, we will consider two groups ($j = A,B$). Equation 4.1 is first estimated separately for both groups using OLS, in order to retrieve the estimated coefficients for each group.

The difference between the outcome variable for the two groups can be expressed as the difference between the wage equation of group A and group B.

$$\overline{lnY_A} - \overline{lnY_B} = (\hat{\alpha}_A + \hat{\beta}_A \bar{X}_A) - (\hat{\alpha}_B + \hat{\beta}_B \bar{X}_B) \qquad \text{[Eq. 4.2]}$$

$$\overline{lnY_A} - \overline{lnY_B} = \underbrace{\bar{X}_B\left(\hat{\beta}_A \bar{X}_A - \hat{\beta}_B \bar{X}_B\right) + (\hat{\alpha}_A - \hat{\alpha}_B)}_{\text{Unexplained Component}} + \underbrace{(\bar{X}_A - \bar{X}_B)\, \hat{\beta}_A}_{\text{Explained Component}} \qquad \text{[Eq. 4.3]}$$

This equation can be further transformed by choosing one of the groups or a combination of both as a reference for the earnings function. In order to ensure the relevance of this choice, many studies have proposed alternative methods than the one originally proposed by Oaxaca (1973). In the standard method, the reference group is the non-discriminated one except for cases of suspected nepotism in which the reference group should be the non-discriminated one. Alternative methods propose to weigh the coefficients by the share of each group in the total population (Reimers 1983) or to use the weighted average coefficients of the two groups as the reference coefficients (Cotton 1988). Neumark (1988) proposes to retrieve the coefficients from a pooled equation that excludes the group variable. Fortin (2008) demonstrates that the Neumark decomposition overstates the unexplained part of the wage gap. She shows that this issue can be resolved by including the group variable in the pooled regression from which it is possible to retrieve the coefficients $\hat{\beta}_p$ yielding the following decomposition equation:

$$\overline{lnY_A} - \overline{lnY_B} = (\bar{X}_A - \bar{X}_B)\, \hat{\beta}_p + [\bar{X}_A\left(\hat{\beta}_A - \hat{\beta}_p\right) + \ (\hat{\beta}_A - \hat{\beta}_p)] - (\bar{X}_B\left(\hat{\beta}_B - \hat{\beta}_p\right) + \hat{\beta}_B - \hat{\beta}_p)]$$

[Eq. 4.4]

Fortin shows that the unexplained component of the wage gap is the sum of the advantage of group A: $\bar{X}_A\left(\hat{\beta}_A - \hat{\beta}_p\right)$ and the disadvantage of group B: $\bar{X}_B\left(\hat{\beta}_B - \hat{\beta}_p\right)$.

This framework allows evacuating arbitrary choices of the reference earnings equations, especially in the case where multiple groups are being compared. Using this method, we can include the selection terms retrieved from the selection equation, $(\theta selP_{1,}\ \theta selP_{2,}\ \theta selP_3)$ as shown in equation 4.5.

$$\overline{lnY_A} - \overline{lnY_B} = (\bar{X}_A - \bar{X}_B)\,\hat{\beta}_p + [\bar{X}_A(\hat{\beta}_A - \hat{\beta}_p) + \widehat{(\beta}_A - \hat{\beta}_p)] - (\bar{X}_B(\hat{\beta}_B - \hat{\beta}_p) + \hat{\beta}_B - \hat{\beta}_p)] +$$
$$(\theta selP_{1A} + \theta selP_{2A} + \theta selP_{3A}) - (\theta selP_{1B} + \theta selP_{2B} + \theta selP_{3B}) \quad \text{[Eq. 4.5]}$$

3.3. A non-parametric decomposition method: the Ñopo matching method

The second step of our analysis allows us to verify whether the lack of common support creates a bias in the estimation of the size of the explained and unexplained parts of the wage gap. According to Ñopo (2008), the OB decomposition framework is bound to provide biased results if the individuals who are being compared do not have comparable characteristics. He proposes a non-parametric matching method to decompose wage gaps.[85]

Nopo demonstrates how the wage gap can be decomposed into four components in the following way. Consider two groups in a given population (e.g. women and men), group A and group B. Y refers to individual earnings. X is an n-dimensional vector of individual characteristics. $F^A(.)$ and $F^B(.)$ are the conditional cumulative distribution functions of individual characteristics X depending on the group considered. $dF^A(.)$ and $dF^B(.)$ are their corresponding probability measures. $\mu^A(S)$ denotes the probability measure of the set S under the distribution $dF^A(.)$. Therefore, $\mu^A(S) = \int_S \, dF^A(x)$ and $\mu^B(S) = \int_S \, dF^B(x)$. The expected value of earnings conditional of characteristics and groups are denoted by $g^A(.)$ and $g^B(.)$. Therefore, $E[Y|A,X] = g^A$ (X). $E[Y|B,X] = g^B$ (X) and $E[Y|A] = \int_{S^A} g^A(x)dF^A(x)$ and $E[Y|B] = \int_{S^B} g^B(x)dF^B(x)$. S^A and S^B are the supports of the distribution of characteristics for each group.

The income gap between group A and group B can be defined as Equation 4.6.

$$\Delta \equiv E[Y|A] - E[Y|B] \qquad \text{[Eq. 4.6]}$$

Which is equivalent to:

$$\Delta = \int_{S^A} g^A(x)dF^A(x) - \int_{S^B} g^B(x)dF^B(x) \qquad \text{[Eq. 4.7]}$$

[85] Note that this method has been implemented to analyze the gender wage gap by Agrawal and Vanneman (2014) in an unpublished working paper to analyze the gender wage gap in the first wave of IHDS data.

Ñopo shows that after a few computations this expression is equivalent to the following expression:

$$\Delta = \left[\int_{\overline{S^B}} g^A(x)\frac{dF^A(x)}{\mu^A(\overline{S^B})} - \int_{S^B} g^A(x)\frac{dF^A(x)}{\mu^A(S^B)}\right]\mu^A(\overline{S^B}) + \int_{S^A\cap S^B} g^A(x)\left[\frac{dF^A}{\mu^A(S^B)} - \frac{dF^B}{\mu^B(S^M)}\right]$$

$$+ \int_{S^A\cap S^B}[g^A(x) - g^B(x)]\,\frac{dF^B(x)}{\mu^B(S^A)} + \left[\int_{S^A} g^B(x)\frac{dF^B(x)}{\mu^B(S^A)} - \int_{\overline{S^A}} g^B(x)\frac{dF^B(x)}{\mu^B(\overline{S^A})}\right]\mu^B(\overline{S^A})$$

[Eq. 4.8]

In short, the wage gap Δ can be written as follows:

$$\Delta = \Delta_{\boldsymbol{A}} + \Delta_{\boldsymbol{X}} + \Delta_{\boldsymbol{0}} + \Delta_{\boldsymbol{B}} \quad \text{[Eq. 4.9]}$$

And reorganized as:

$$\boxed{\Delta = \Delta_{\boldsymbol{X}} + \Delta_{\boldsymbol{0}} + \Delta_{\boldsymbol{A}} + \Delta_{\boldsymbol{B}} \quad \textbf{[Eq. 4.10]}}$$

$\Delta_X = \int_{S^A\cap S^B} g^A(x)\left[\frac{dF^A}{\mu^A(S^B)} - \frac{dF^B}{\mu^B(S^M)}\right]$. It is the part of the wage gap that can be explained by differences in the distribution of characteristics of individuals from groups A and B over the common support. It is equivalent to the *explained component of OB decompositions*.

$\Delta_0 = \int_{S^A\cap S^B}[g^A(x) - g^B(x)]\,\frac{dF^B(x)}{\mu^B(S^A)}$. It is the part of the wage gap that cannot be explained by differences in individual characteristics and is attributed to potential discrimination or unobservable characteristics (productivity, motivation etc.). It refers to the *unexplained component of OB decompositions.*

$\Delta_A = \left[\int_{\overline{S^B}} g^A(x)\frac{dF^A(x)}{\mu^A(\overline{S^B})} - \int_{S^B} g^A(x)\frac{dF^A(x)}{\mu^A(S^B)}\right]\mu^A(\overline{S^B})$. It is the difference between expected wages from group A individuals out of the common support and the expected wages of group A individuals in the common support. It refers to the part of the wage gap which is due to differences between two groups of individuals from group A, those whose characteristics can be matched to group B and those whose characteristics cannot be matched. Therefore, it is the part of the gap that would disappear if all individuals from group A would match individuals from groups B or if the wage of the unmatched individuals from group A would on average be the same as those of matched individuals from group B.

$\Delta_B = \left[\int_{S^A} g^B(x)\frac{dF^B(x)}{\mu^B(S^A)} - \int_{\overline{S^A}} g^B(x)\frac{dF^B(x)}{\mu^B(\overline{S^A})}\right] \mu^B(\overline{S^A})$. This part of the gap is the one explained by differences in characteristics between two groups inside of the B group, those who have characteristics that can be matched to group A characteristics and those who cannot.

The matching algorithm allows taking into consideration the distributional characteristics of one of the groups using a replacement method. In the case of gender discrimination, Ñopo considers that the distributional characteristic of the smallest group (we assume it is group B) should be preserved. Therefore the matching method is the following.

1. Selection of one group B individual from the sample (without replacement)
2. Select all the group A individuals who have the same characteristics than the individual chosen in step 1.
3. Construct a synthetic individual whose characteristics are equal to the average of all of the individuals chosen in step 2. Moreover, match this synthetic individual to the group B individual chosen in step 1.
4. Put the synthetic group A individuals and the chosen group B individual in their respective groups, respectively: matched group A and matched group B.
5. Repeat steps 1 through 4 until it exhausts the original group B sample.

We estimate wage gaps across the gender, religion and caste groups. The results indicate whether a lack of common support is an issue in the parametric decomposition method.

3.4. A description of wage gaps along the distribution

Ñopo's nonparametric method provides results at the mean level, but the samples generated to estimate the wage gap can be used to analyze its distribution. We use these comparisons as a descriptive tool to observe how the wage gap is distributed and the potential existence of sticky floors (a high wage gap at the beginning of the distribution) and glass ceilings (a high wage gap at the end of the distribution). In order to provide these observations, we compute the percentiles of wages for group A and B, in the whole sample and the matched sample. Using these percentiles, we plot the distribution of the wage gap across the percentiles, thus providing a visual tool to analyze wage gaps.

Following this description, we implement decompositions across various quantiles to see the extent to which the differentials can be attributable to individual characteristics or potential discrimination. Two alternative approaches exist in the analysis of wage gaps along the distribution: the Machado-Mata-Melly (MMM) decomposition framework and the Recentered Influence Function Regression (RIF-Reg) decomposition framework (Fortin, Lemieux, and Firpo 2011a). We briefly describe these methodologies in Appendix 4.1. In these decompositions, the sample selection bias is not controlled for. The analysis of the source of wage gaps along the distribution would require a separate study since the selection terms computed from the selection correction method cannot be included in the estimations. Töpfer (2017) shows that a specific non-parametric method can be used to correct the selection bias in a RIF-reg approach. For these reasons, we only briefly comment on the results from these decompositions.

4. Data and descriptive statistics

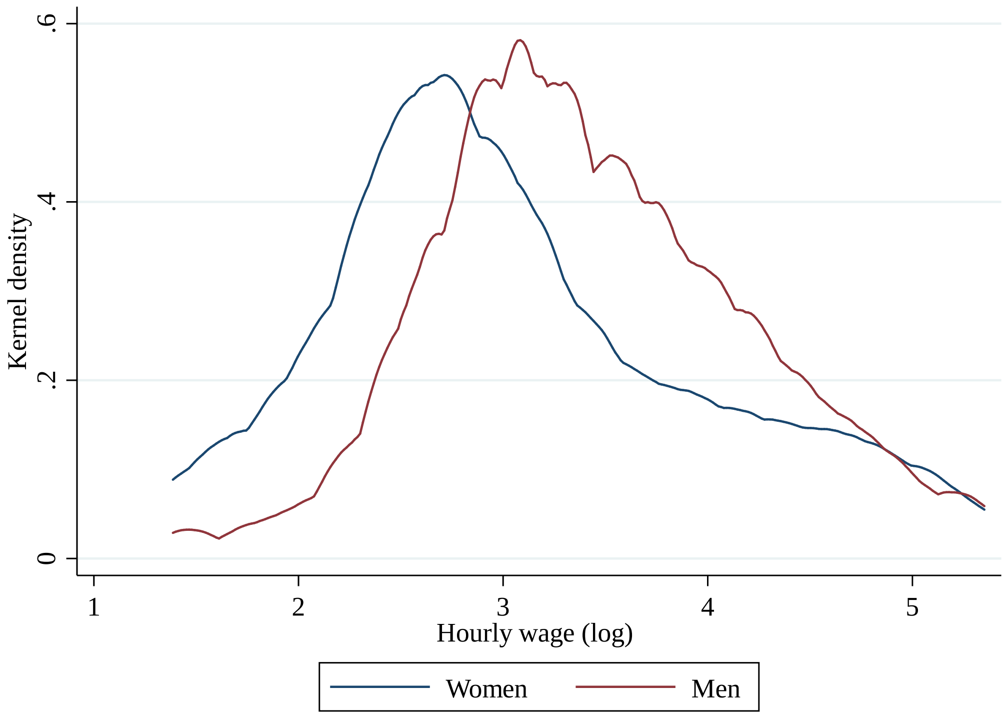
We use the second wave of the IHDS data in the analysis. The sample of interest is restricted to active occupied individuals from urban areas who are between 15 and 65 years old (both values included). In order to detect potential employer discrimination, we conduct our analysis exclusively on salaried workers, thus excluding self-employed individuals. The latter group may also face discrimination, but the mechanisms may be different since they are likely to suffer from customer or supplier discrimination. We exclude individuals who are below the first and over the 99th percentile of the wage distribution. The final sample is composed of 14,661 wage earners.

In this study, we analyze the differentials in the natural logarithm of hourly earnings for salaried workers, which we also address as wages. The list of explanatory variables used in the decomposition analyses is presented in the descriptive statistics table in Appendix 4.2.[86] We use the same variables as those used in chapter 3 except for one additional variable which is English

[86] Note that in Mincer earnings functions, a variable indicating the level of experience should also ideally be included. In our analysis age and its squared measure are the only proxies of experience. We do not include a potential experience variable (the difference between the age and the number of years before and during school) because it tends to be biased. It is a good proxy for men but not for minority groups or for women who have a higher likelihood of professional activity interruption (Altonji and Blank 1999; Nordman and Roubaud 2009). This approximation would not be relevant for this study.

ability. This variable distinguishes individuals who do not have any English proficiency (coded 0) from those who have a partial proficiency or more (coded 1).

Figure 4.1. Kernel density graph of log hourly wages by gender

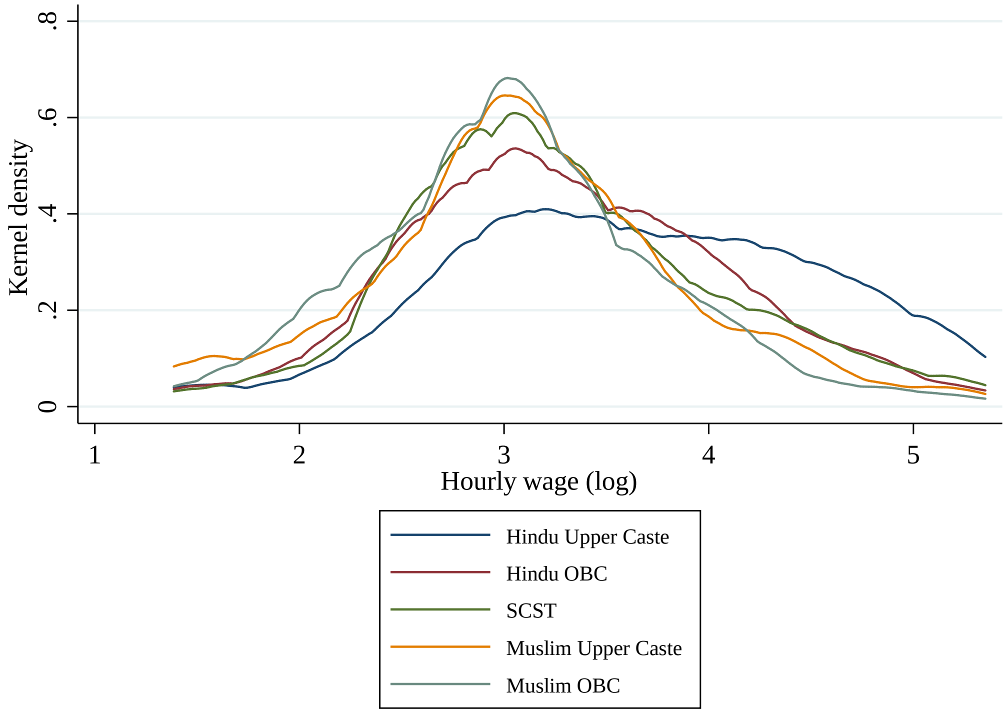


Source: Author's calculations from IHDS (2011-12)

Figure 4.1. shows the kernel density functions of women's and men's log wages. Women earn less than men with a more right-skewed density function. Women earn on average 35.17 INR per hour compared to 42.47 INR for men. The descriptive statistics in Appendix 4.2. show that working women are more frequently uneducated and from the SCST groups. Although they only represent 22.23% of the working-age population, 31.47% of active women and 26.37% of active men are SCSTs. However, a second profile of active women is visible in the statistics with more tertiary educated women in the labor market relative to men, and more women with an SSC first-class level.

Figure 4.2 shows that concerning socio-religious groups, Hindu Upper Castes have a distribution that is more left-skewed than the other groups. The other groups have similar distributions regarding skewness. Note that the large right-tail of the Hindu OBC group suggest the existence of higher wages for this group compared to other groups except for the Hindu Upper Caste one.

Figure 4.2. Kernel density graph of log hourly wages by religion and caste

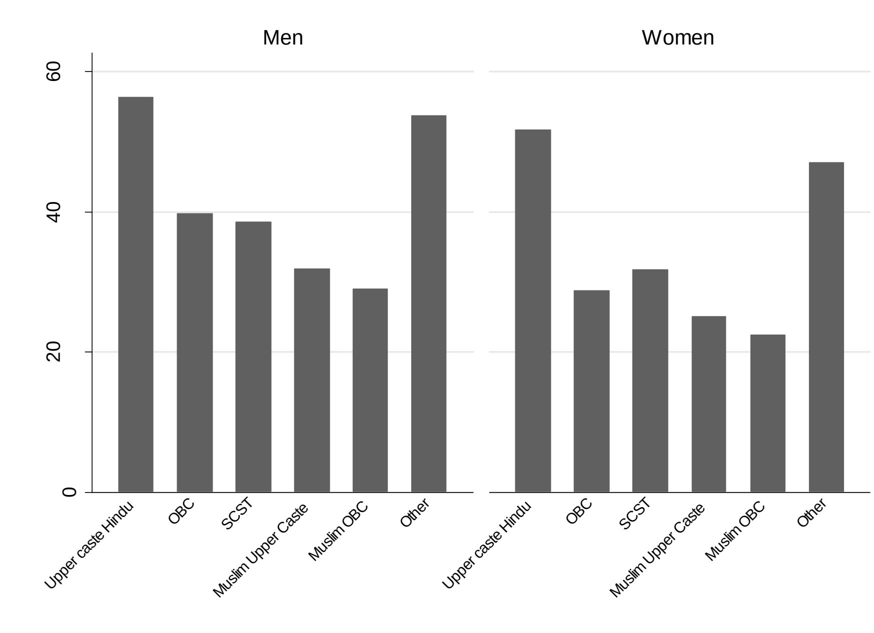


Source: Author's calculations from IHDS (2011-12)

Appendix 4.3. shows that the Hindu Upper caste group beneficiates from higher shares in secondary and higher education. The share of individuals from this group who were ranked in the first class of the SSLC examination (25.06%). is also much higher than the other groups (between 12.88% and 5.62%).

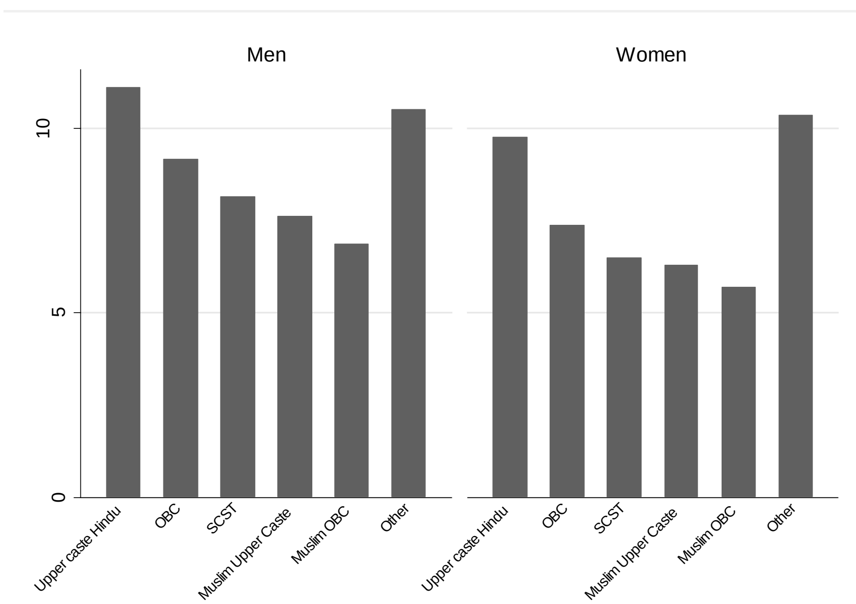
The intersection of gender and socio-religious groups (Figure 4.3) shows that the Muslim group has the lowest average hourly earnings in both male and female samples. We also see that the Hindu Upper Caste group has the highest average hourly earnings for both gender groups. All distributions are skewed to the left. 50% of Muslim women earn 12.50 rupees or less whereas 50% of Hindu Upper Caste women earn 27.40 rupees or less. The intersection of gender and caste shows that regarding educational attainment (Figure 4.4), the most unschooled groups seem to be Hindu OBC, SCST and Muslim women. On the other hand, the groups with the largest share of highly educated individuals are Hindu Upper Caste women and men. These results already show substantial endowment gaps between groups and these should affect incomes.

Figure 4.3. Hourly earnings by gender and caste [87]



Source: Author's calculations from IHDS (2011-12)

Figure 4.4. Education by gender and caste



Source: Author's calculations from IHDS (2011-12)

[87] Note that the means are different from wage decomposition results since we use the natural logarithm of the earnings variable in the estimations and the predicted means are geometric means.

5. Results

In this section, we present the results from the wage decompositions by gender (5.1) and socio-religious groups (5.2.). In a third section (5.3.), we explore whether the decomposition results indicate forms of joint discrimination.

5.1. Decompositions of earnings differentials by gender

5.1.1. Parametric decompositions at the mean

Table 4.1. Parametric gender decomposition results

	Without selection correction		With selection correction	
	Coefficient	**Percent of the gap**	**Coefficient**	**Percent of the gap**
Male reference wage structure				
Characteristics	0.009 (0.011)	**2.6%**	0.347*** (0.087)	**99.4%**
Coefficients	0.340** (0.015)	**97.4%**	0.003 (0.087)	**0.6%**
Female reference wage structure				
Characteristics	0.024 (0.017)	**6.8%**	-0.161 (0.185)	**-46%**
Coefficients	0.325*** (0.019)	**93.2%**	0.510*** (0.186)	**146%**
Pooled reference wage structure				
Characteristics	0.058*** (0.012)	**16.6%**	0.335*** (0.019)	**96.0%**
Unexplained component				
Due to male advantage	0.058*** (0.003)	**16.6%**	0.003*** (0.000)	**0.3%**
Due to female disadvantage	0.233*** (0.011)	**66.8%**	0.012*** (0.002)	**3.7%**
Total wage gap		0.349		

Source: Author's calculations from IHDS (2011-12)
Note: N=14,661 Bootstrapped standard errors in parentheses (500 replications) *** p<0.01, ** p<0.05, * p<0.1

Table 4.1. shows the parametric decompositions of the gender wage gap. The results fluctuate considerably depending on the reference wage structure and the sample selection bias correction. Without selection correction, most of the wage gap is due to a coefficient (or unexplained) effect. When the male wage structure is considered as the reference, 97.4% of the wage gap is unexplained and is not due to differentials in observable productivity characteristics. When the female wage structure is considered as the reference, there is a slight decline in the coefficient effect by

approximately four percentage points. The decompositions using the pooled wage structure show that the characteristics effect (explained component of the wage gap) is about 16.6%. This result is in line with previous studies that do not account for the sample selection bias. For instance, in a study of the gender wage gap among regular salaried workers (i.e. excluding casual workers), Deshpande, Goel, and Khanna (2017) find a discriminatory component of 88.2% to 111.1% between 1999-2000 and 2009-10 respectively.[88] Conversely, our results provide better control for productive ability with the SSC rank variable.

When selection into employment is controlled for, the results change substantially. If men's earnings function is chosen as the reference, the wage gap is almost exclusively due to a coefficient effect. The difference in the probability of working between men and women is largely responsible for the wage gap. The pooled decomposition estimation shows quite similar results with a 96% explained component and a 4% unexplained one[89]. Adding the selection terms has two effects on the results. First, it corrects the values of the other coefficients. Second, the selection term themselves are viewed as variables and contribute to explained and unexplained shares of the wage gap. Keeping this in mind, the results do not imply an absence of wage discrimination in the labor market. The detailed decomposition results provided in Appendix 4.1 show that the selection terms highly contribute to reshaping the shares of the explained and unexplained component, but the magnitude of the other variables, namely productive characteristics variables remain the same. The results show that the differential in the probability of working is the main explanation of the male-female wage gap. Note that even when controlling for the sample selection bias 4% of the wage gap remains unexplained, mostly due to a disadvantage of the female group which points to the existence of minor wage discrimination. Furthermore, selection into employment, which is the main cause of the wage gap, might be caused by other forms of discrimination such as employment discrimination.

---

[88] Note that in their wage function, they include occupations and union membership which are not included in our estimations

[89] Note that our findings are very different from the study of Kingdon and Unni (2001) in Madhya Pradesh. After controlling for selection into employment with a Heckman method, they found that between 18% and 55.7% of the gender wage gap remained unexplained.

5.1.2. Nonparametric decomposition results

The method developed by Ñopo consists in comparing “comparable” individuals. Its main purpose in this study is to observe whether Oaxaca-Blinder decomposition results are prone to a lack of common support bias. It is only possible to include dummy variables in the matching process, which implies that some variables such as the squared-age cannot be added in the Ñopo specification. With a more precise matching, the common support gets smaller. It is therefore important to find the right balance between the number of matching variables and the share of the population in the common support. Also, an interesting feature of this method is that the comparison of the results from different matching models enables us to identify which variables artificially inflate Δ_x and Δ_0 because of the lack of common support. For instance, a common support that changes when we add a variable indicating occupational type shows that occupational segregation contributes to the wage gap. It leads some workers to not find counterparts to be matched with, and consequently, the wage gap between these individuals and the rest of the group would fall into the explained component Δ_x and the unexplained component Δ_0 of an OB-type decomposition. This method allows us to identify the actual reasons behind wage differentials to better isolate the unexplained effect. The sum of the four additive components Δ_0, Δ_x, Δ_{men}, Δ_{women} is equal to the total wage gap (Δ) of the whole sample. The explained part of the income gap Δ_x is the part of the gap that is due to a different distribution of characteristics among men and women in the common support. Δ_{men} is the part of the wage gap that is due to the wage differential between *men who are comparable to women* (inside of the common support) and *men who are not* (outside of the common support). Δ_{women} is the part of the wage gap that is due to the wage differential between *women who are comparable to men* and *women who are not*. The results are expressed as the share of the logged female wage.

We compare the results from the five sets of matching variables. The incremental addition of the variables allows better identification of the source of the wage gap. All variables are categorical variables which are included in the specification as dummy variables. We also computed dummy variables for age.

Table 4.2. Nonparametric matching models to decompose the gender wage gap

Model 1	Age and highest educational attainment
Model 2	Age, highest educational attainment and SSC class (none, 1, 2 or 3)
Model 3	Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 categories)
Model 4	Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 categories), casual occupation
Model 5	Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 categories), casual occupation, type of occupation (8 skill categories)
Model 6	Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 categories), casual occupation, type of occupation (8 skill categories) and religion caste group

Source: Author

Table 4.3. Nonparametric decompositions of the gender wage gap

	Coefficient	Percent of the gap	Percent of matched men	Percent of matched Women	Wage gap	Wage gap in the matched sample
Model 1	***age and highest educational attainment***					
Δ	0.115[90]	100%	97.68	100	0.345***	0.358***
Δ_0 (std. error)	0.104 (0.007)	90.4%				
Δ_X	0.014	12.2%				
Δ_{men}	-0.003	**(-)2.6%**				
Δ_{women}	0.000	**0%**				
Model 2	*Age, highest educational attainment and **SSC class (none, 1, 2 or 3)***					
Δ	0.115	100	94.6	99.5	0.345***	0.357***
Δ_0 (std. error)	0.110 (0.007)	95.7%				
Δ_X	0.005	4.3%				
Δ_{men}	0.000	**0%**				
Δ_{women}	0.000	**0%**				
Model 3	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), **sector of occupation (5 categories)***					
Δ	0.115	100%	73.63	96.20	0.345***	0.322***
Δ_0 (std. error)	0.105 (0.003)	91.3%				
Δ_X	0.000	0%				
Δ_{men}	0.008	**6.7%**				
Δ_{women}	0.002	**1.7%**				

Table 4.3 continued on next page

[90] This result is expressed as the share of the average female wage. A gender gap of 0.115 therefore equates to a gap of 0.354 in the Oaxaca-Blinder decomposition.

Table 4.3 (continued)

Model 4	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 categories),* ***casual occupation***					
Δ	0.115	100%				
Δ_0	0.112	97.3%				
(std. error)	(0.000)					
Δ_X	0.000	0%	65.31	93.21	0.345***	0.344***
Δ_{men}	0.000	**0%**				
Δ_{women}	0.003	**2.6%**				
Model 5	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 categories), casual occupation,* ***type of occupation (8 skill categories)***					
Δ	0.115	100%				
Δ_0	0.123	107%				
(std. error)	(0.000)					
Δ_X	-0.012	(-)10.4%	35.77	77.53	0.345***	0.345***
Δ_{men}	-0.008	**(-)7%**				
Δ_{women}	0.012	**10.4%**				
Model 6	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 categories), casual occupation, type of occupation and* ***religion and caste group***					
Δ	0.115	100%				
Δ_0	0.127	110.4%				
(std. error)	(0.000)					
Δ_X	-0.000	0%	35.77	77.53	0.345***	0.345***
Δ_{men}	-0.034	**(-)29.5%**				
Δ_{women}	0.022	**19.1%**				

Source: Author's calculations from IHDS (2011-12)
Note: N=14,661. Student Ttest show significant differences between the wages of men and women in the matched sample, and in the whole sample. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

The results presented in Table 4.3 show that the total gender wage gap (Δ) amounts to 11.5% of the mean of women's log wages[91]. The explained components of the most precise models (5 and 6) are very small. We would expect the wage gaps to be smaller among individuals matched by skill level, but the results from Model 5 shows that, on the contrary, the wage gap becomes higher when we add the type of occupation. Furthermore, the nonparametric decomposition results show that the lack of common support potentially affects the results from OB decompositions since the intragroup heterogeneity of the male and female samples are likely to wrongly be considered as a between-group heterogeneity, by contributing to Δ_X and Δ_0. The values of Δ_{men} and Δ_{women} significantly increase when the type of occupation is added (model 5), which confirm the existence

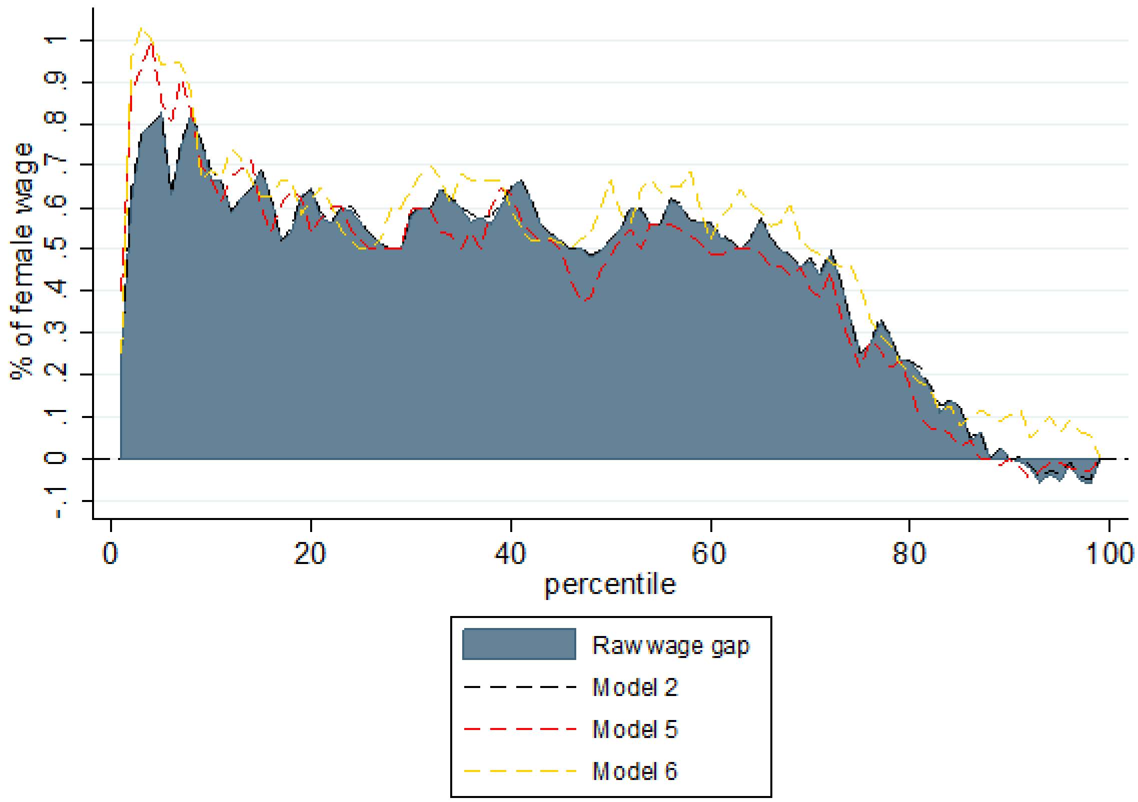
[91] The logged female and male wages being 3.077 and 3.431, the difference between both, 0.355 is 11.5% of the female wage.

of a lack of common support. Note that the negative value of Δ_{men} means that the men who cannot be matched to women earn more than the men who can be matched. Symmetrically, the positive value of Δ_{women} shows that women who can be matched to men earn more than women who cannot. The common support is composed of individuals who are “in the middle”. The men who have no female counterparts in the common support earn more than the men who do and women who have no counterparts in the common support earn less than other women. This result suggests the existence of occupational segregation. Some women have the worst paying jobs in the labor market and men do not have access to these jobs. Conversely, some men have the better paying jobs and women do not have access to these jobs.

5.1.3. The gender wage gap along the distribution

An interesting feature of the non-parametric decomposition performed in the previous section is that it allows seeing the distribution of the wage gap between individuals who are comparable on the basis of the matching variables. Using the matched samples from the Ñopo analysis, Figure 4.5 shows the distribution of the wage gap among the whole sample (Raw wage gap) and the matched samples of models 2,5 and 6. The wage gap is expressed as a percentage of women’s wages.

Figure 4.5. Distribution of the gender wage gap



Source: Author's calculations from IHDS (2011-12)

All the distributions follow the same trend. The gap is very high at the beginning of the distribution (80% for the raw wage gap and up to 100% for models 5 and 6). It remains approximately at 50% between the 20th and 60th percentile. A steep decrease is visible after the 65th percentile and there is a small wage gap or a negative one after the 90th percentile. In a nutshell, the trends describe the existence of an important "sticky floor" phenomenon, an important gap all along the distribution, and an absence of a "glass-ceiling phenomenon."

The results in Appendix 4.5 show the quantile decomposition results (MMM and RIF-reg). Both sets of estimation were conducted using the same variables as for the parametric decompositions, without the sample selection variables. The results clearly show that all along the quantiles, productive characteristics do not explain a large part of the wage gap. It is mostly unexplained all along the distribution suggesting potential discrimination. As mentioned in Section 3.4, the fact that these results do not account for sample selection calls for additional investigations. Moreover, the discrepancy between the actual wage gaps and the predicted wage gaps in the RIF framework show that this method might not be adapted and also calls for additional research.

In their study of the wage gap among regular wage workers using an MMM decomposition method., Deshpande, Goel, and Khanna (2017) also find the existence of a sticky floor which is mostly due to potential discrimination (73% of the gap at the 10^{th} percentile). They provide several explanations for this situation. The discrimination women face at the beginning of the distribution is probably statistical discrimination. Men are perceived to be more reliable and less likely to exit the labor market for household responsibilities than women. Another explanation is that since the nature of occupations at the two ends of the distribution is different, women who are at the highest percentiles of the distribution are more likely to know their rights engage in legal actions in cases of discrimination. Moreover, occupational segregation may contribute to a higher gap at the lower percentiles. The results we provided point to the existence of substantial occupational segregation which explains the gender wage gap at the mean but also at the lower percentiles of the distribution.

5.2. Decompositions of earnings differentials by socio-religious groups

5.2.1. Parametric decomposition results

Table 4.4 presents the results from the parametric wage decompositions by religion and caste. Compared to the findings concerning gender, adding the correction of the sample selection does not affect the results concerning religion and caste as much. In other words, the severity of the non-random allocation of workers across unemployment, self-employment and salaried employment situations between caste and gender groups is not as critical as between gender groups. Nonetheless, the comparison of the two sets of results show a few notable effects of selection which we will describe.

Since the study compares the situation between five groups, one potentially advantaged group (Hindu Upper Castes) and four potentially disadvantaged groups (Hindu OBC, SCST, Muslim Upper Castes, and Muslim OBC), we decompose the wage differentials between each group and the rest of the population. Other studies in the case of India usually compare SCSTs to the rest of the population or Hindu Upper Castes to the rest of the population. In this perspective, providing the results for each group is a contribution of our study.

Approximately 11% of the wage gap between Hindu Upper Castes and the rest of the population remains unexplained after controlling for personal characteristics. The differential points toward the existence of nepotism since 8.1% of the gap is due to an advantage of Hindu Upper Castes. The remaining 2.7% is due to the disadvantage of the other groups. These results suggest the existence of nepotism in the urban labor market, and to a lesser extent, the existence of discrimination against non-Hindu Upper Caste workers. With the sample selection correction, the unexplained share increases to 13% and is almost exclusively due to Hindu Upper Caste advantage.

The Hindu OBC group is the only disadvantaged group which faces potential discrimination, accounting for approximately 58% of the wage gap. When the selection correction is applied, the unexplained component decreases (42%). In both cases, this differential is due to a disadvantage of the Hindu OBC group. For all of the other groups, the results show that the differences are mostly due to differentials in personal characteristics.

The results point to the coexistence of Hindu Upper Caste nepotism and discrimination against Hindu OBCs. The disadvantage of the other groups (SCSTs and both Muslim groups) are mostly due to differentials in personal characteristics.

Table 4.4. Parametric religion and caste decomposition results

Reference group	**Hindu Upper Castes**		**Hindu OBC**		**SCST**		**Muslim Upper Caste**		**Muslim OBC**	
	Without selectivity correction	**With selectivity correction**	**Without selectivity correction**	**With selectivity correction**	**Without selectivity correction**	**With selectivity correction**	**Without selectivity correction**	**With selectivity correction**	**Without selectivity correction**	**With selectivity correction**
	Wage gap 0.367***		**Wage gap -0.081*****		**Wage gap -0.111**		**Wage gap -0.256**		**Wage gap -0.321**	
Reference group wage structure										
Characteristics	0.426*** (0.020)	0.395*** (0.028)	0.009 (0.165)	-0.267*** (0.019)	-0.111*** (0.012)	0.084 (0.054)	-0.129*** (0.037)	-0.122** (0.058)	-0.231*** (0.044)	-0.228*** (0.732)
Coefficients	-0.059*** (0.021)	-0.030 (0.028)	-0.091*** (0.018)	-0.029*** (0.006)	0.001 (0.014)	-0.196*** (0.055)	-0.127*** (0.040)	-0.134** (0.064)	-0.090* (0.018)	-0.095 (0.074)
Other group wage structure										
Characteristics	0.277*** (0.012)	0.260*** (0.012)	-0.035*** (0.012)	-0.045*** (0.011)	-0.143*** (0.012)	-0.085*** (0.025)	-0.229*** (0.020)	-0.240*** (0.022)	-0.307*** (0.016)	-0.327*** (0.018)
Coefficients	0.090*** (0.016)	0.104*** (0.015)	-0.045*** (0.010)	-0.040*** (0.014)	0.033*** (0.012)	-0.026 (0.025)	-0.027 (0.025)	-0.016 (0.026)	-0.014 (0.019)	0.003 (0.021)
Pooled wage structure										
Characteristics	0.327*** (0.011) **89.1%**	0.319*** (0.012) **86.9%**	-0.034*** (0.100) **(-)42.0%**	-0.047*** (0.011) **(-)58.0%**	-0.123*** (0.011) **(-)110.8%**	-0.101*** (0.014) **(-)91.0%**	-0.229*** (0.019) **(-)89.5%**	-0.238*** (0.021) **(-)93.0%**	-0.302*** (0.016) **(-)94.1%**	-0.320*** (0.016) **(-)99.6%**
Unexplained component										
→ Due to reference group	0.030*** (0.009) **8.1%**	0.034*** (0.008) **9.3%**	-0.033*** (0.007) **(-)40.7%**	-0.026*** (0.007) **(-)32.1%**	0.010 (0.007) **9.0%**	-0.008 (0.004) **(-)7.2%**	-0.025 (0.021) **(-)9.7%**	-0.017 (0.021) **(-)6.6%**	-0.018 (0.016) **(-)5.6%**	-0.001 (0.015) **(-)0.4%**
→ Due to other groups	0.010*** (0.009) **2.7%**	0.003*** (0.003) **0.8%**	-0.015*** (0.003) **(-)18.5%**	-0.012*** (0.003) **(-)14.8%**	0.004 (0.003) **0.4%**	-0.003 (0.002) **(-)2.7%**	-0.002 (0.001) **(-)0.8%**	-0.001 (0.002) **(-)0.4%**	-0.001 (0.001) **(-)0.3%**	-0.000 (0.001) **(-)0%**

Source: Author's calculations from IHDS (2011-12)

Note: N=14,661 Bootstrapped standard errors in parentheses (500 replications) *** p<0.01, ** p<0.05, * p<0.1

5.2.2. Non-parametric decompositions

The matching variables in the nonparametric decomposition by religion and caste group are presented in Table 4.5 The decomposition results are presented in Table 4.6. for Hindu Upper Castes compared to the other groups and in Appendix 4.5 for the other configurations.

Table 4.5. Nonparametric matching models to decompose the religion and caste wage gap

Model 1	Age and highest educational attainment
Model 2	Age, highest educational attainment and SSC class (none, 1, 2 or 3)
Model 3	Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 categories)
Model 4	Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 categories), casual occupation
Model 5	Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 categories), casual occupation, type of occupation (8 skill categories)
Model 6	Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 categories), casual occupation, type of occupation and gender

Source: Author

Table 4.6. Nonparametric decomposition of socio-religious wage gap (Hindu Upper Caste compared to other groups)

	Hindu Upper Caste compared to other groups					
	Coefficient	Percent of the gap	% matched in ref. group	% of matched in rest of sample	Wage gap	Wage gap in the matched sample
Model 1	***age and highest educational attainment***					
Δ	0.112	100 %	100	96.48	0.368***	0.352***
Δ_0 (std. error)	-0.003 (0.002)	-2.7 %				
Δ_X	0.110	98.2 %				
Δ_A	**0.000**	0 %				
Δ_B	**0.005**	4.4 %				
Model 2	*Age, highest educational attainment and* ***SSC class (none, 1, 2 or 3)***					
Δ	0.112	100 %	99.33	95.05	0.368***	0.351***
Δ_0 (std. error)	-0.008 (0.007)	-7.1 %				
Δ_X	0.115	102.7 %				
Δ_A	**0.000**	0 %				
Δ_B	**0.004**	3.6 %				
Model 3	*Age, highest educational attainment, SSC class (none, 1, 2 or 3),* ***sector of occupation (5 categories)***					
Δ	0.112	100 %	95.21	76.06	0.368***	0.322***
Δ_0 (std. error)	0.004 (0.000)	3.6 %				
Δ_X	0.094	83.9 %				
Δ_A	**0.011**	9.8 %				
Δ_B	**0.002**	1.8 %				

Table 4.6 continued on next page

Table 4.6 (continued)

Model 4	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 categories), **casual occupation***					
Δ	0.112	100 %				
Δ_0	-0.001 (0.000)	-0.9 %				
Δ_X	0.092	82.1 %	90.44	66.99	0.368***	0.299***
Δ_A	**0.002**	1.8 %				
Δ_B	**0.019**	16.9 %				
Model 5	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 categories), casual occupation, **type of occupation (8 skill categories)***					
Δ	0.112	100 %				
Δ_0	-0.003 (0.000)	2..7%				
Δ_X	0.078	69.6%	70.81	49.62	0.368***	0.265***
Δ_A	**0.005**	4.46%				
Δ_B	**0.027**	24.1%				
Model 6	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 categories), casual occupation, type of occupation and **gender***					
Δ	0.112	100 %				
Δ_0	0.007 (0.000)	6.25%%				
Δ_X	0.062	55.4%	35.77	77.53	0.368***	0.227***
Δ_A	**0.002**	1.8%				
Δ_B	**0.040**	35.7%				

Source: Author's calculations from IHDS (2011-12)
Note: N=14,661 Standard errors in parentheses (500 replications) *** p<0.01, ** p<0.05, * p<0.1

The results show that the more precise the matching, the lower the wage gap is, which suggests that there is a common support issue (note that this is not the case for the gender wage gap). The different components of the wage gap show that when individuals are matched on the basis of educational attainment and ability (Model 2), the wage gap is almost exclusively due to a differential in characteristics. When we include variables indicating the type of occupation,Δ_A and Δ_B increase, showing that there are individuals from both groups who cannot be matched on the basis of occupational characteristics. The difference between Models 3 and 4 is the addition of casual-regular occupation as a matching variable. In Model 3, Δ_B accounts for approximately 2% of the wage gap and in Model 4, Δ_B accounts for 17% of the wage gap. Therefore, approximately 15% of the wage differential between Hindu Upper Castes and other groups is due to the fact that the other groups do not occupy the same occupations in terms of casual and regular employment.

The non-parametric decomposition results suggest that occupational segregation accounts for a considerable part of the wage gap between Hindu Upper Castes and the other groups. The rest of the gap is mostly due to differentials in characteristics. In comparison to the parametric decomposition results, the nonparametric results confirm that the potential wage discrimination is less than 10%. The lack of common support, in this case, falls into the explained component of the parametric wage decompositions.

Concerning the other groups, Appendix 4.5 confirms the existence of potential discrimination against Hindu OBCs. The shares of Δ_A and Δ_B also suggest the existence of occupational segregation for all of the groups. The negative sign and the high contribution of Δ_B in all Model 5 and 6 decomposition show that each caste group apart from the Hindu Upper Caste occupies specific occupational niches in which they earn more than individuals from their group who are matched.

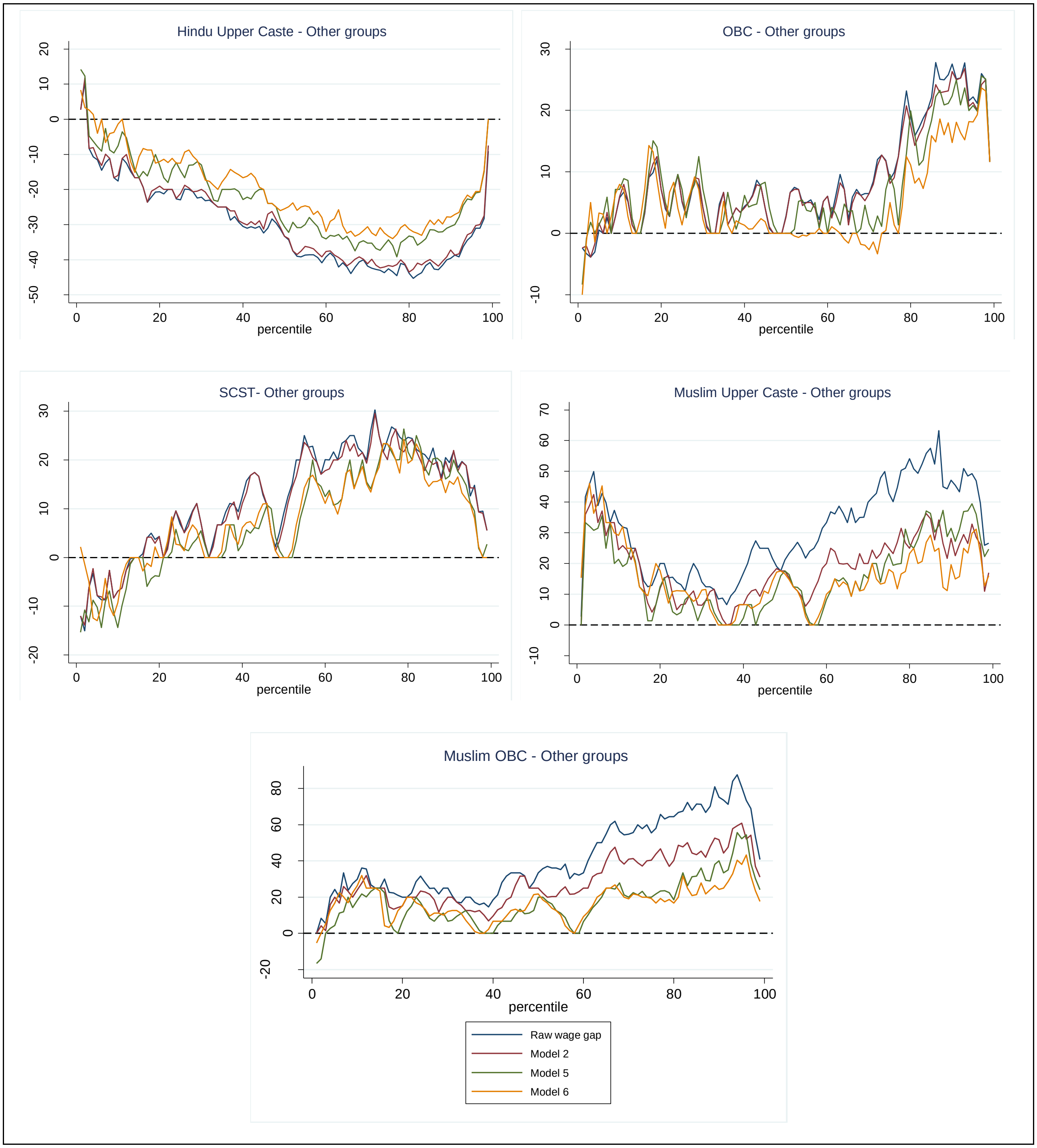
5.2.3. The wage gap along the distribution

The graphs in Figure 4.6 show that the wage gap expressed in % of the Hindu Upper caste wage between the latter group and the rest of the population takes a U-shape. The negative sign in the y-axis means that Hindu Upper Castes earn more than the rest of the population. The gap is the highest between the 60th and 80th percentile. The wage gap between Hindu OBCs and the other groups is higher at the end of the distribution, especially after the 80th percentile, suggesting that this group faces a glass-ceiling. Concerning SCSTs, the wage gap is particularly visible after the median but considerably drops after the 90th percentile. Both groups (Hindu OBC and SCST) have a similar wage gap magnitude, which is up to 30% of their own wage. By contrast, the wage gap between Muslims and the rest of the population is much higher as it can attain 60% for Muslim Upper Castes and 80% for Muslim OBCs.

In all cases, the wage gap diminishes in the matched samples. In the graphs concerning Muslims, the space between the raw wage gap and the gap from Model 2 (matching on the basis of education level) shows that a large part of the wage gap, especially at the end of the distribution, is due to the low wage of Muslim individuals who cannot be matched in terms in term of education and SSC rank.

The quantile decomposition estimation results presented in Appendix 4.6. show that the wage differential between Hindu Upper Caste and the other groups is mostly due to a characteristic differential.

Figure 4.6. Religion and Caste wage gaps along the distribution



Source: Author's calculations from IHDS (2011-12)

5.3. An insight on joint discrimination

Banerjee et al. (2009) find that in formal sector call-center jobs, female SCSTs do not face discrimination whereas males suffer from lower call-back rates than other groups. They do not discuss this gender difference further in their study. However, it is possible that both types of stigmas (gender on one side and caste or religion on the other) have a compensating effect in the labor market. Conversely, both factors may add up to create groups that are particularly vulnerable. This section proposes to observe the potential joint discrimination between both groups.[92]

First, the parametric decomposition by gender shows that the wage gaps are higher among women. This result is in line with the literature considering the fact that working women in India are mostly driven to the labor market by necessity (i.e. working poor women), but there is also a group of qualified women in the labor market. The decomposition of wages between Hindu Upper Castes and the rest of the population shows the wage gap (in favor of Hindu Upper Castes) is mostly due to a characteristic differential. Conversely, Hindu Upper Caste men seem to benefit from an unexplained advantage which can be considered as nepotism and other groups seem to suffer from a disadvantage. In the female subsample, potential discrimination is only visible in the Muslim Upper Caste. For all the other groups, the unexplained component of the wage gap is small and non-significant.

[92] A first version of this study analyzed gender gaps inside of caste and religion groups as well as caste and religion gaps inside of gender groups. We chose to present only the results from the latter group since the analysis of labor market segmentation in chapter 3 suggested that there is important selection into occupations in the case of India.

Table 4.7. Religion and Caste wage gaps along the distribution

Decomposition	Hindu Upper Castes vs. Others		Hindu OBC vs. Others		SCST vs. Others		Muslim Upper Caste vs. Others		Muslim OBC vs. Others	
	Without selectivity correction	**With selectivity correction**	**Without selectivity correction**	**With selectivity correction**	**Without selectivity correction**	**With selectivity correction**	**Without selectivity correction**	**With selectivity correction**	**Without selectivity correction**	**With selectivity correction**
Female subsample										
Total wage gap	**0.475*****	**0.473*****	**-0.182*****	**-0.186*****	**-0.106*****	**-0.110*****	**-0.397*****	**-0.402*****	**-0.437*****	**0.441*****
Reference wage structure										
Characteristics	0.527*** (0.060)	0.595*** (0.091)	0.001 (0.061)	-0.077 (0.074)	-0.096*** (0.035)	0.123 (0.131)	-0.297* (0.161)	-0.310 (0.244)	-0.438*** (0.156)	-0.421 (0.257)
Coefficients	-0.052 (0.054)	-0.121 (0.089)	-0.183*** (0.064)	-0.109 (0.154)	-0.009 (0.040)	-0.233* (0.133)	-0.100 (0.158)	-0.092 (0.245)	0.001 (0.157)	-0.020 (0.266)
Other group wage structure										
Characteristics	0.388*** (0.031)	0.365*** (0.036)	-0.159*** (0.028)	-0.180*** (0.032)	-0.180*** (0.028)	-0.133** (0.056)	-0.218*** (0.062)	-0.240*** (0.070)	-0.434*** (0.047)	-0.477*** (0.054)
Coefficients	0.087** (0.042)	0.109** (0.044)	-0.023 (0.029)	-0.006 (0.032)	0.074** (0.031)	0.023 (0.056)	-0.180** (0.078)	-0.163** (0.083)	-0.003 (0.057)	0.036 (0.063)
Weighted wage structure										
Characteristics	0.446*** (0.032)	0.442*** (0.034)	-0.147*** (0.026)	-0.174*** (0.028)	-0.128*** (0.027)	-0.092*** (0.033)	-0.255*** (0.058)	-0.277*** (0.070)	-0.436*** (0.046)	-0.476*** (0.050)
Unexplained component										
→ Due to reference group	0.022 (0.021)	0.024 (0.020)	-0.025 (0.016)	-0.009 (0.015)	0.014 (0.017)	-0.013 (0.010)	-0.137** (0.064)	-0.120* (0.070)	-0.001 (0.050)	0.033 (0.047)
→ Due to other groups	0.007 (0.006)	0.008 (0.006)	-0.011 (0.008)	-0.003 (0.007)	0.008 (0.008)	-0.005 (0.005)	-0.005** (0.002)	-0.005* (0.003)	-0.000 (0.003)	0.001 (0.003)
Male subsample										
Total wage gap	**0.335*****	**0.332*****	**-0.052*****	**-0.056*****	**-0.092*****	**-0.094*****	**-0.264*****	**-0.262*****	**-0.333*****	**-0.335*****
Reference wage structure										
Characteristics	0.385*** (0.020)	0.323*** (0.039)	0.024 (0.017)	0.008 (0.021)	-0.096*** (0.013)	0.150* (0.087)	-0.185*** (0.033)	-0.251*** (0.063)	-0.247*** (0.051)	-0.228*** (0.087)
Coefficients	-0.050** (0.022)	0.008 (0.038)	-0.076*** (0.020)	-0.064*** (0.023)	0.004 (0.016)	-0.244*** (0.087)	-0.079** (0.038)	-0.012 (0.068)	-0.087 (0.055)	-0.107 (0.091)
Other group wage structure										
Characteristics	0.243*** (0.012)	0.218*** (0.013)	-0.005 (0.012)	-0.003 (0.012)	-0.114*** (0.011)	-0.074** (0.032)	-0.255*** (0.020)	-0.266*** (0.020)	-0.302*** (0.017)	-0.321*** (0.020)
Coefficients	0.092*** (0.017)	0.113*** (0.018)	-0.047*** (0.015)	-0.054*** (0.016)	0.021* (0.012)	-0.019 (0.032)	-0.008 (0.025)	0.003 (0.025)	-0.031 (0.021)	-0.014 (0.023)
Weighted wage structure										
Characteristics	0.290*** (0.012)	0.281*** (0.013)	-0.006 (0.010)	-0.010 (0.010)	-0.104*** (0.011)	-0.090*** (0.015)	-0.252*** (0.019)	-0.260*** (0.019)	-0.300*** (0.016)	-0.318*** (0.018)
Unexplained component										
→ Due to reference group	0.033*** (0.010)	0.038*** (0.008)	-0.032*** (0.008)	-0.032*** (0.008)	0.008 (0.008)	-0.003 (0.004)	-0.011 (0.021)	-0.002 (0.020)	-0.031* (0.017)	-0.016 (0.016)
→ Due to other groups	0.011*** (0.003)	0.013*** (0.003)	-0.014*** (0.004)	-0.015*** (0.004)	0.003 (0.003)	-0.001 (0.001)	-0.000 (0.001)	-0.000 (0.001)	-0.002* (0.002)	-0.001 (0.001)

Source: Author's calculations from IHDS (2011-12)

Note: N=14,661 Bootstrapped standard errors in parentheses (500 replications) *** p<0.01, ** p<0.05, * p<0.1

In the nonparametric decompositions, it is possible to observe the existence of joint discrimination by comparing the models 5 to 6. Indeed, models 5 contain information on education, productivity-related factors and occupation. In the models 6 we add the religion and caste variables or the gender variable.

The decomposition of the gender wage gap in Table 4.7 shows that the values of Δ_A and Δ_B increase significantly when the religion and caste variable is added in the matching process. This indicates that the interaction of gender and religion or caste is linked to occupational segregation and that the gender wage gap is largely composed of individuals who cannot be matched on the basis of the type of occupation and religion and caste.

Box 5. Perception of discrimination in Ranipet's leather industry

During the interviews in Ranipet, we asked the workers whether they felt that their religion, caste and gender affected the course of their professional lives.

A few workers mentioned that they benefitted from a religious network to find employment in the leather industry. These workers were all from the Muslim community. The question regarding caste was generally not answered. Concerning gender, about one-third of the women spoke about the choice of employment in the leather factory, which was an "obvious choice for women here (in Ranipet)."

Following this question, we asked the workers whether they had ever experienced a form of work-related discrimination. Most of the times, we had to precise which type of discrimination (gender, caste and religion). The unanimous answer was "no" concerning gender discrimination. Concerning religion, workers pointed out the existence of discrimination against Hindus in Muslim factories taking the form of unequal bonuses for festivals and although there was no information sharing on wages between the Muslim and Hindu community working in the same factory, the Hindu workers suspected they were paid less. Furthermore, some workers pointed out that although Tamils were employable in Muslim tanneries, they could only work in companies in which the spoken language was Tamil and not Urdu. Indeed, the main language of Tamil Nadu is Tamil but members of the Muslim community also speak Urdu which is generally not understood by the non-Muslim individuals in Ranipet. Furthermore, Hindu companies do not hire Muslims according to one of the workers.

Two arguments can shed light on the unanimous answer of women regarding the absence of discrimination. First, there might be no work-related discrimination in Ranipet. As shown in box 4, women in the leather factories benefit from working conditions that are relatively better than their spouses who work in tanneries. Nevertheless, this claim ignores the structural dimension of gender discrimination which causes women to interrupt their careers and engage in more casual forms of labor such as the new forms of "*putting-out systems*" that exist in Ranipet. Furthermore, the discourse of women points out that they decide to engage in specific occupations based on "what is usual" rather than a choice. In this case, there might not be employment discrimination since there are no occurrences where a woman would choose to do a job destined for men. A second explanation is that individuals may not perceive discrimination even though it exists. Klumpp and Su (2013) show that detecting the presence and absence of discrimination is difficult for individuals. False positives (i.e. impression of prejudice) and false negatives (i.e. being discriminated against without perceiving it) may be present in the labor market. They show that individuals will only observe outcomes and not the process to determine whether they face discrimination. In this case, since men and women do not compete for the same occupations, it seems logical that the perception of discrimination is less likely.

The fact that some workers pointed out the existence of a bonus differential between Hindus and Muslims is very interesting as it may reflect a form of nepotism towards Muslims in the specific case in which the employer is Muslim. It may also be a case of false positive, in which the worker feels discriminated against although there is no evidence (concerning wages) that there is unequal pay. Religion-based occupational segregation can be caused by employer discrimination or by the self-selection of workers in occupations where they are less likely to face discrimination.

Note that these questions were generally asked at the end of the interviews which has two advantages. After describing their careers, it was easier for the interviewees to adopt a holistic perspective. Moreover, it allowed ensuring that a minimum level of trust and contact was established.

Source: Author

Conclusion and discussion

This chapter analyzes the nature of gender and socioreligious wage gaps in the Urban Indian labor market. By implementing several methodologies, we provide clear evidence that educational variables do not contribute to the gender wage gap and we provide suggestive evidence that occupational segregation is the main explanation of wage gaps. Allocation of workers into different labor market statuses (non-participant, self-employment and salaried employment) and in different types of occupations (casual or regular and by skill-level) contribute to the gender wage gap. The results show that women occupy the lower paying jobs, especially at the lower percentiles of the wage distribution.

Regarding socio-religious variables, our results show the existence of nepotism in favor of Hindu Upper Castes and the existence of discrimination against Hindu OBCs. The wage gap between Hindu Upper Castes and other groups is partly due to employment segregation. The wage gap between SCSTs and the other groups is mostly due to differentials in characteristics, which is also the case for Muslim workers. The findings contribute to the part of the literature that considers that wage gaps between socio-religious groups are due to the unequal allocation of workers in different segments rather than pure wage discrimination (Ito 2009)

Some covariates in the parametric decompositions might be correlated to the error term, which causes the zero conditional mean assumption of OLS to fail. However, this does not necessarily affect the results as long as the correlation is the same for each group (Fortin, Lemieux, and Firpo 2011). One omitted variable that probably impacts the results and requires further investigation is the intermittence of occupations for women. In this study, we use age and its square value as a proxy for age, which is probably insufficient. An improvement for further studies will be to investigate the correct way to proxy experience.

General Conclusion

This dissertation explores the existence and the extent of horizontal inequalities based on gender, religion and caste in India. Focusing on four specific research questions, we combine quantitative methodologies with insights from a field study in Ranipet to provide an empirically-grounded picture of group disadvantage in the Indian labor market.

The **first chapter** analyzes the links between labor market exclusion and group disadvantage. After describing horizontal inequality in premarket factors (i.e. education and health) the first section shows the direct and indirect correlations between gender-, religion- and caste-based group membership and labor market exclusion. The probability of labor market exclusion for socio-religious groups is mostly mediated by education. However, being a woman significantly increases the probability of labor market exclusion, regardless of the level of education. The second section of this chapter addresses the reverse association between labor market exclusion and the perpetuation of gender disparity across generations. Mothers' labor market participation, more specifically the fact that she is allowed to work, contributes to broadening the gender gap in education. Girls who have higher scores are in households in which female labor is stigmatized. This result leads to the conclusion that they will probably remain outside of the labor market when they reach adulthood. We find no effect of school-related time use on the gender gap in education. A mother's full-time labor market participation significantly diminishes the hours spent doing homework for her sons and daughters, suggesting transmission of household chores to children. We suppose that for female labor to narrow the gender gap through a motivational effect on girls, female work should be empowering and not solely related to subsistence.

The **second chapter** analyzes two dimensions of medium-run labor market mobility between 2005 and 2011-12. In order to detect occupational mobility, we measure transitions across casual and regular employment, industries and the skill level requirement of occupations. We also compare patterns of relative hourly earnings mobility. The results show that women are more immobile than men concerning occupational mobility, but they do not have significantly different relative earnings mobility. Hindu Upper Castes have higher rates of movers into higher skilled employment and the services sector, but they have a lower hourly earnings mobility compared to all groups, suggesting a process of "catching-up" We estimate the determinants of mobility by considering multiple potential sources of bias. Education level is an important determinant of upward occupational mobility. Nonetheless, after controlling for personal characteristics such as education or ability, significant differentials are visible among the socio-religious groups. Compared to Hindu Upper Castes, SCSTs have a lower likelihood of mobility toward better quality occupations (i.e. regular job or higher skill requirement). Muslims also have significantly lower chances of transitioning into regular employment. Belonging to a specific caste group does not significantly affect the chances of relative earnings mobility except at the top of distribution where Muslim Upper Castes face significantly lower chances of mobility. Two trends affect women in opposing ways. Their lower levels of education limit their transitions into higher skilled jobs, but for equivalent levels of initial earnings, education and ability among other factors, they benefit from a higher likelihood of upward skill mobility than men. The results show that the trend of catching-up for disadvantaged group goes through education. The religion and caste groups with lower socioeconomic status benefit from less occupational mobility and less earnings mobility than Hindu Upper Castes. Despite upward occupational mobility, women do not benefit from a significant relative change in the distribution. These conclusions, which are based, on the measurement of hourly earnings do not rule out that access to more hours of work or public transfers can be a source of economic mobility.

The **third chapter** analyzes the existence of labor market segmentation in a predominantly informal labor market. Using a semi-parametric method to detect the presence of segments, we show that the household business sector is better represented by a homogenous structure whereas a segmented one best represents the salaried sector. The detected segmentation points to strong gender segregation of the labor market. Indeed, women are part of a distinct labor market segment with a lower average wage than the two other segments combined. Men are divided into two segments: an Upper Male Segment and a Lower Male Segment. We show that the Female Segment and the Lower Male Segment constitute traps inside the labor market. Informality being predominant in urban India, the better-quality jobs are concentrated in the Upper Male Segment, but we find no evidence of a clear formal/informal divide. Barriers of access to this segment are present not only for women but also for specific socioreligious groups as all non-Upper-Hindu Caste groups are less present in Upper Male Segment than in Lower Male Segment.

The **fourth chapter** analyzes the sources of wage gaps by comparing parametric and non-parametric decomposition results. We find that the gender wage gap is not due to pure wage discrimination. The selection effect into occupation and segregation into different types of occupations are the main source of wage differentials between women and men. The wage gap between Hindu Upper Castes and the rest of the population is partly due to nepotism, to discrimination and to endowment differentials. The selection and segregation across occupation only contribute to the wage gap between this group and the rest of the population, but to a lesser extent than for gender. Hindu OBC is the group that suffers the most from this discrimination, as the wage gap of the other disadvantaged groups compared to the rest of the sample is almost exclusively due to endowment differentials.

The main findings from our research highlight the following points.

The first form of substantial horizontal inequality encountered in the labor market is the differential in premarket factors. Educational policies that promote equal access to education are therefore crucial to ensure that individuals have the same endowments before entering the labor market. Affirmative action ensuring reserved seats in universities can for instance be extended. Indeed, if the labor market were to become less segmented, structural group disparity would contribute to the persistence of horizontal inequality.

The analysis of the gender wage and employment gap is an important contribution of this paper. The latter is hard to detect with survey data, which is why our conclusions can only mostly pertain to potential discrimination. Discrimination can occur at many levels, and our findings suggest that a male bias in households can increase the educational gap, especially in households where mothers work full-time. We also find that a minimal presence of pure wage discrimination in the labor market, which is relevant considering the strong occupational segregation on the lines of gender. The wage gaps between men and women exist because women are in lower-wage occupations compared to men. This nuance may reveal lower reservation wages for women, as a supply-side characteristic. It may also reveal a demand-side characteristic as employers believe that women’s work is less productive because of statistical discrimination (Phelps 1972). A third possibility is that it may be the consequence of the long-run interactions between both mechanisms (Bergmann 1974). In this case, women’s lower wage become a social norm, leading them to “choose” specific types of low-wage employment to conform to these norms. This third possibility implies that the distinction between discrimination and potential discrimination no longer holds. Concerning socio-religious groups, the results also suggest forms of nepotism and discrimination. If segregation is the cause of wage inequality, formalization in the long-run may contribute to lowering wage gaps. Indeed, formal enterprises have to abide by non-discriminatory laws which is likely to lead to an increase in female labor. The coexistence of different male and female wages in a formal setup increases the odds of negotiation. However, this may only contribute to a partial diminution of the wage gap. Indeed, gender wage gaps tend to persist despite development and despite formalization, as it can be witnessed in many developed countries. Furthermore, the fact that social norms act as an institution to determine labor market outcomes implies that policies that encourage access to better occupations are likely to be inefficient in the short-term. In this case, it is essential to ensure that women who are segregated in low-quality jobs benefit from short-term public policy such as universal social protection.

Horizontal-inequality among socio-religious group differs depending on the group of interest. If it is evident that the non-discriminated group is the Hindu Upper Caste, it does not necessarily imply that the other groups are discriminated against. The results from the segmentation and wage gap analyses point out that segregation of specific groups in segments of the labor market contributes to the gap between Hindu Upper Castes and the other groups. Furthermore, the existence of nepotism and discrimination against OBCs is also visible. SCSTs and Muslims, who are substantially economically disadvantaged in the labor market, suffer mostly from disparities in education which contribute to lower earnings. Occupational mobility can have a equalizing nature if it is associated with an increase in income, however our findings show no significant leaps in hourly earnings for the disadvantaged group between 2005 and 2011-12.

Finally, the results also show interesting associations between gender and socio-religious groups. Gender and caste interact in the determination of women's participation in the labor market. It also matters inside of the female segment of the labor market in which Muslims OBC women earn significantly less than other women. Interestingly, nepotism does not exist between Hindu Upper Caste women and the rest of the population. These results echo the literature on joint discrimination by suggesting that cumulating identities overlap and may increase or decrease stigma (Figart 1997).

The findings from this thesis could benefit from several extensions and improvements. First, the systematic comparison of the results from the two waves of data could provide interesting information concerning the medium-run dynamics of our findings. Furthermore, conducting a State-specific analysis may prove useful since the way gender and socioreligious groups are perceived is not homogeneous across the Indian territory. Moreover, the analysis of the intergenerational effect of female labor market participation could be completed in many ways, such as distinguishing secluded women from those who are not allowed to work for other reasons. Finally, a considerable improvement of our findings would be to integrate a more disaggregated level of the socio-religious group. A recent study from Joshi, Kochhar, and Rao (2017) point out that caste may not be the most relevant desegregation level when analyzing gender inequality as the focus on *jatis* reveals considerable nuances.

Appendix

General Appendix

Appendix 0.1. Methodology for the qualitative study

A. Map of Ranipet

Leather industrial towns in Vellore district (Tamil Nadu)

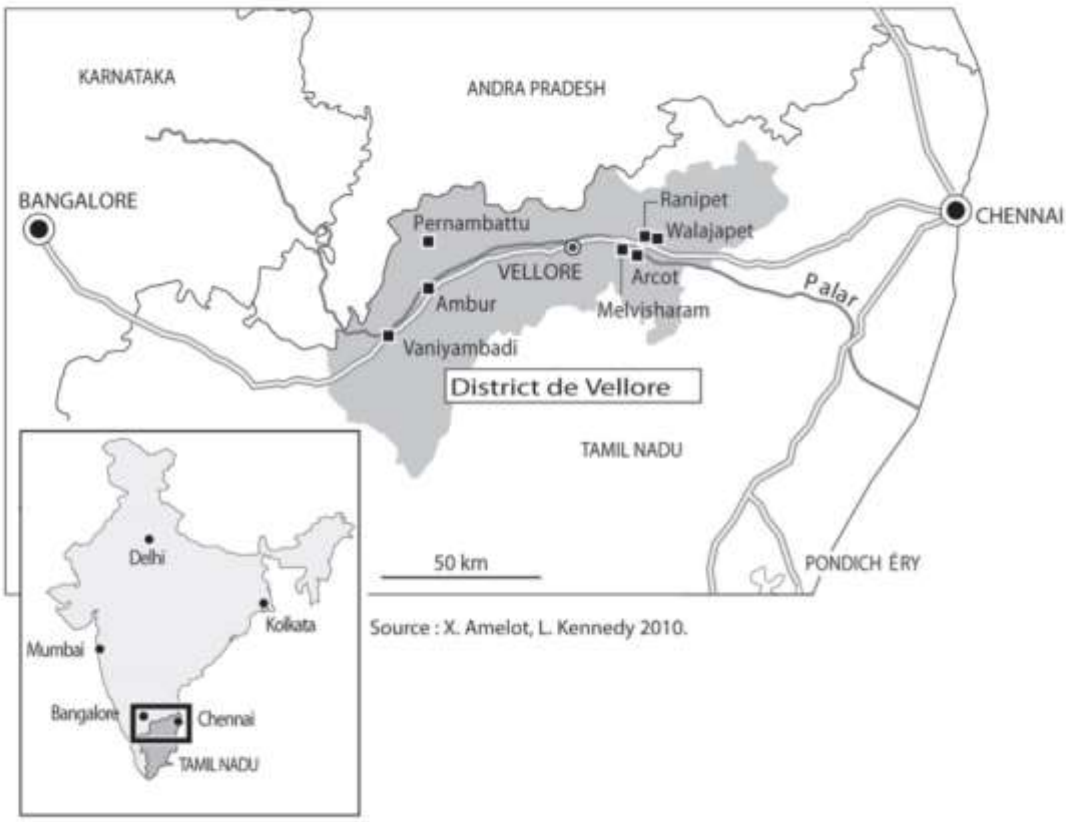


Fig. 1: Leather industrial towns in Vellore district (Tamil Nadu)

Source: Marius and Venkatasubramanian (2017)

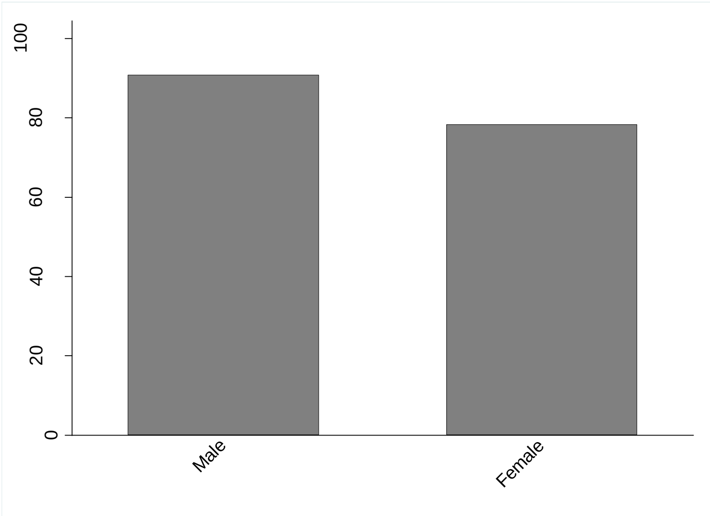
B. Survey and Analysis of Qualitative data

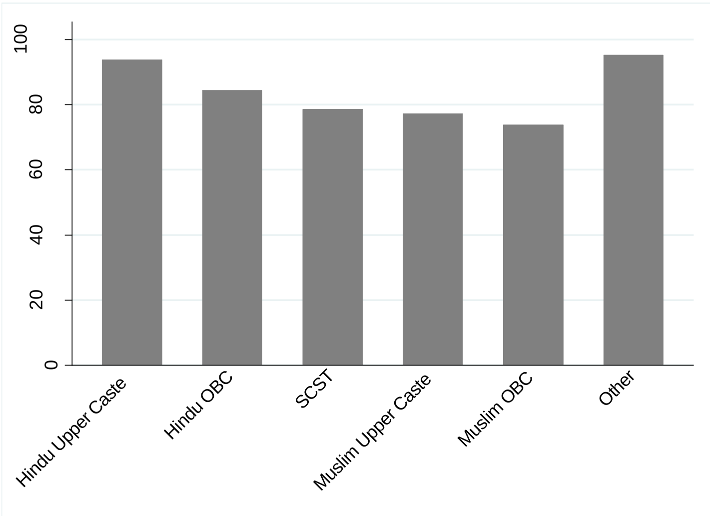
The qualitative study aimed to retrieve information on the labor market conditions of the leather industry workers in Ranipet. Consequently, most of the individuals we interviewed were workers, that we interviewed alone or with their family members. Moreover, in order to understand the relevance of our findings, we conducted additional interviews with employees from the Employment provisions Fund and a health center. Most of the workers were interviewed in their homes, before or after their day of work. We also interviewed a few workers and an employer from a tannery (which was a small production unit of fewer than 10 workers). In total, approximately 50 to 60 individuals were interviewed.

The interviews lasted from approximately 20 minutes to 1.5 hours. After a brief presentation of our research interests, we asked the interviewees to describe their occupations, professional history, regular tasks and career perspectives. We also addressed their parents' occupations and what were the expectations for their children's occupations. Moreover, we discussed labor conditions, employment relations, and unionization. At the end of the discussion, the workers were asked how they considered themselves concerning labor (occupation and/or income) in comparison to other caste groups and the other gender group. One of the challenges of the data collection was to address the potential existence of gender and caste stigmas on different labor market outcomes. The answers given at the time allowed us to view the perception of agents on the labor market. Therefore, our aim is not specifically to detect whether the perceptions of agents reflect reality, but rather to understand how the way they perceive their situation in the labor market can affect their future decisions. The qualitative database was analyzed and encoded using NVIVO software that allowed classifying elements of the discourse in different categories and subcategories (nodes) in order to identify elements that interviewees have in common and elements that stand out.

Appendix to Chapter 1

Appendix 1.1. Adult literacy rate by gender, religion and caste groups





Source: Author's calculations from IHDS (2011-12)

Appendix 1.2. Column and row percentages of literacy rates among gender, religion and caste groups

		Literacy					
		Column percentage (%)			**Row percentage (%)**		
	N	No	Yes	Total	No	Yes	Total
Gender							
Male	34,856	37.5	53.6	50.2	15.6	84.4	100
Female	34,543	62.5	46.4	49.8	26.2	73.8	100
Total	69,399	100	100	100	20.9	79.1	100
	Pearson chi2(1) = 1188.905 P-value = 0.000						
Religion/Caste							
Hindu Upper Caste	18,095	14.9	29.3	26.3	11.7	88.3	100
Hindu OBC	21,056	30.8	30.6	30.7	20.9	79.1	100
SC-ST	15,228	27.7	20.7	22.2	25.9	74.1	100
Muslim Upper Caste	4,782	9.1	6.4	7	27.1	72.9	100
Muslim OBC	7,123	15.8	8.9	10.4	31.7	68.3	100
Other	2,403	1.7	4	3.5	10.3	89.7	100
Total	68,687	100	100	100	20.8	79.2	100
	Pearson chi2(5) = 1935.264 P-value = 0.000						

Source: Author's calculations from IHDS (2011-12)

Appendix 1.3: Column and row percentages of Gender, Religion and Caste group distribution across education levels

		Education Level													
		Column percentage (%)							Row percentage (%)						
	N	None	Kindergarten	Primary	Middle	Secondary	Higher	Total	None	Kindergarten	Primary	Middle	Secondary	Higher	Total
Gender															
Male	25904	28.3	49.6	45.8	53.2	55.6	56.2	49.5	9.8	6.1	6.1	26.4	19.2	32.5	100
Female	26445	71.7	50.4	54.2	46.8	44.4	43.8	50.5	24.4	6	7.1	22.8	15	24.7	100
Total	52349	100	100	100	100	100	100	100	17.2	6	6.6	24.6	17.1	28.6	100
Pearson chi2(5) = 1227.641 P-value = 0.000															
Religion/Caste															
Hindu Upper Caste	14405	17.3	18.5	23.3	32.3	42.2	27.8	12.9	7.9	3.7	4.4	20.5	19.9	43.6	100
Hindu OBC	15952	32.1	31.6	32.3	32	27.9	30.8	31.5	17.5	6.3	6.7	25.7	17.8	26	100
SCST	11255	24.3	25.8	23.4	18.7	15.9	21.7	29.6	23.4	6.7	7.8	26.4	14.7	21	100
Muslim Upper Caste	3448	9	7.4	7.6	5.2	4.3	6.7	9.6	24.9	8.2	7.3	27.9	13.4	18.4	100
Muslim OBC	4824	14	13.7	10.1	7.1	4.5	9.3	15.1	27.9	9.1	9.7	26.5	13	13.9	100
Other	1945	3.3	3.1	3.3	4.7	5.3	3.8	1.2	5.5	5.3	5.3	21.5	21.5	40.8	100
Total	51829	100	100	100	100	100	100	100	17.1	6	6.6	24.5	17.1	28.7	100
Pearson chi2(25) = 4804.237 P-value = 0.000															

Source: Author's calculations from IHDS (2011-12)

Appendix 1.4. Column and row percentages of Gender, Religion and Caste group distribution across the short-term illness variable

	Short-term Illness										
		Column percentages (%)					Row percentages (%)				
	N	Not ill	Ill for less than one week	Ill for less than two weeks	Ill for more than two weeks	Total	Not ill	Ill for less than one week	Ill for less than two weeks	Ill for more than two weeks	Total
Gender											
Male	25941	51.2	38.1	36	38	49.5	90.4	7.3	1.3	1	100
Female	26472	48.8	61.9	64	62	50.5	84.5	11.6	2.3	1.7	100
Total	52413	100	100	100	100	100	87.4	9.5	1.8	1.3	100
	Pearson chi2(3) = 416.0979 P-value = 0.000										
Religion/Caste											
Hindu Upper Caste	14411	27.9	27.2	25.5	25.4	27.8	87.8	9.3	1.7	1.2	100
Hindu OBC	15983	30.9	30.3	28.2	28.8	30.8	87.7	9.4	1.6	1.3	100
SC-ST	11262	21.5	22.4	25.2	23.4	21.7	86.7	9.8	2.1	1.4	100
Muslim Upper Caste	3452	6.7	6.2	6.9	7.9	6.7	87.7	8.9	1.9	1.6	100
Muslim OBC	4834	9.2	9.7	11.7	11.6	9.3	86.2	9.9	2.3	1.7	100
Other	1947	3.7	4.2	2.6	3	3.8	87.1	10.6	1.2	1.1	100
Total	51892	100	100	100	100	100	87.3	9.5	1.8	1.3	100
	Pearson chi2(15) = 36.9120 P-value = 0.001										

Source: Author’s calculations from IHDS (2011-12)

Appendix 1.5. Column and row percentages of Gender, Religion and Caste group distribution across long-term illness variable

	Long-term hospitalization								
	Column percentages (%)					Row percentages (%)			
	N	Never hospitalized	Hospitalized for less than one month	Hospitalized for more than one month	Total	Never hospitalized	Hospitalized for less than one month	Hospitalized for more than one month	Total
Gender									
Male	3168	40.8	47.5	55.7	42.4	74.4	24.4	1.2	100
Female	4299	59.2	52.5	44.3	57.6	79.4	19.9	0.7	100
Total	7467	100	100	100	100	77.3	21.8	0.9	100
Pearson chi2(2) = 27.9801 P-value = 0.000									
Caste/Religion Group									
Hindu Upper Caste	2219	31.5	25.7	14.5	30	80.9	18.6	0.5	100
Hindu OBC	2187	29	31.6	37.7	29.6	75.6	23.2	1.2	100
SC-ST	1388	18	21.2	27.5	18.8	74	24.6	1.4	100
Muslim Upper Caste	518	7.3	6.2	2.9	7	80.3	19.3	0.4	100
Muslim OBC	684	8.9	10.6	11.6	9.3	74	24.9	1.2	100
Other	392	5.4	4.8	5.8	5.3	79.3	19.6	1	100
Total	7388	100	100	100	100	77.3	21.8	0.9	100
Pearson chi2(10) = 43.1552 P-value = 0.000									

Source: Author's calculations from IHDS (2011-12)

Appendix 1.6. Multinomial logit estimation results for full sample

	Labor Market Participation (Reference group: Full-time worker)	
	Labor Market Participation (Reference group: Full-time worker)	**Labor Market Participation (Reference group: Full-time worker)**
Female	3.785***	1.461***
	(0.039)	(0.028)
Hindu OBC	-0.190***	0.158**
	(0.043)	(0.073)
Hindu SC-ST	-0.355***	-0.056
	(0.043)	(0.051)
Muslim upper caste	0.098	-0.192**
	(0.076)	(0.085)
Muslim OBC	0.274***	0.063
	(0.079)	(0.090)
Other groups	-0.122	0.066
	(0.097)	(0.156)
Age	-0.308***	-0.133***
	(0.008)	(0.007)
Age squared	0.004***	0.002***
	(0.000)	(0.000)
Primary school	-0.165	0.043
	(0.129)	(0.095)
Middle school	0.117	0.083
	(0.157)	(0.100)
Secondary school	0.082	-0.153
	(0.127)	(0.092)
Higher education	0.350***	-0.236**
	(0.114)	(0.114)
Literacy	0.160	-0.402***
	(0.117)	(0.099)
Ill for less than one week	0.135	-0.109
	(0.121)	(0.114)
Ill for less than two weeks	0.034	-0.018
	(0.041)	(0.055)
Ill for more than two weeks	-0.370***	-0.358***
	(0.084)	(0.066)
Married	-0.145***	-0.062
	(0.044)	(0.043)
Number of female children (<15 years)	0.355***	0.370**
	(0.130)	(0.163)
Number of male children (<15 years)	-0.008	0.180
	(0.103)	(0.129)

Region control		Yes
Constant	3.156***	0.850***
	(0.201)	(0.153)
Observations (Total 44781)		44781

Source: Author's calculations from IHDS (2011-12)
Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix 1.7. Treatment and outcome estimation results

Tables A. General Score (outcome and treatment models)

	General score Girls			General score Boys		
	Outcome models					
	Not allowed to work and not working	**Allowed to work and not working**	**Working**	**Not allowed to work and not working**	**Allowed to work and not working**	**Working**
Average treatment effect	N.a.	-0.541** (0.222)	-0.516** (0.219)	N.a.	0.354** (0.165)	0.317* (0.163)
Age	0.0636*	0.0852***	0.0800***	0.0928***	0.0465***	0.0957***
	(0.0343)	(0.0161)	(0.0156)	(0.0299)	(0.0155)	(0.0142)
Education kindergarten	0.448*	0.471***	0.612***	0.662***	0.280***	0.521***
	(0.246)	(0.116)	(0.150)	(0.200)	(0.0914)	(0.103)
Education primary	0.597**	0.614***	0.733***	0.689***	0.433***	0.629***
	(0.250)	(0.120)	(0.155)	(0.214)	(0.0963)	(0.106)
Education lower secondary	0.570**	0.539***	0.685***	0.710***	0.463***	0.620***
	(0.259)	(0.128)	(0.160)	(0.222)	(0.101)	(0.110)
Mother’s age	-0.0110**	-0.0101***	-0.00855***	-0.00446	-0.00417	-0.00467*
	(0.00522)	(0.00325)	(0.00299)	(0.00546)	(0.00270)	(0.00250)
School distance	0.00414	0.00773	-0.00160	0.00989**	0.00548*	0.00380
	(0.00801)	(0.00566)	(0.00447)	(0.00469)	(0.00302)	(0.00231)
Number of sisters	-0.0802**	-0.0396**	-0.0990***	-0.0368	-0.0535***	-0.0299*
	(0.0363)	(0.0155)	(0.0172)	(0.0323)	(0.0170)	(0.0158)
Number of brothers	-0.0517	-0.0364*	-0.0799***	-0.0456	-0.0293	-0.0111
	(0.0422)	(0.0203)	(0.0195)	(0.0423)	(0.0184)	(0.0164)
Number of days ill	8.34e-05	-0.000724	0.00570	-0.0153*	-0.000177	-0.00143
	(0.0102)	(0.00492)	(0.00566)	(0.00831)	(0.00458)	(0.00482)
Household head full-time work	-0.0752	0.0918	0.0887	0.0151	0.0198	0.0559
	(0.0724)	(0.0592)	(0.0732)	(0.109)	(0.0455)	(0.0624)
Household head part-time work	-0.0596	0.0779	0.148**	-0.00448	0.0406	0.0691
	(0.0649)	(0.0554)	(0.0719)	(0.0993)	(0.0426)	(0.0617)
Highest Female Education in household	0.0156**	0.0164***	0.0241***	0.0165**	0.0129***	0.0182***
	(0.00629)	(0.00423)	(0.00408)	(0.00680)	(0.00350)	(0.00335)
Highest Male Education in Household	0.0219***	0.0190***	0.0103**	0.0130*	0.0129***	0.0227***
	(0.00776)	(0.00442)	(0.00407)	(0.00742)	(0.00363)	(0.00363)
Hindu OBC	0.0153	0.000831	0.00489	0.0712	-0.00617	0.0165

	(0.0638)	(0.0384)	(0.0374)	(0.0643)	(0.0316)	(0.0369)
SCST	0.0192	-0.133***	-0.0970**	-0.117	-0.0851**	-0.0803**
	(0.0602)	(0.0451)	(0.0420)	(0.0852)	(0.0340)	(0.0363)
Muslim Upper Caste	-0.157	0.0371	0.0805	-0.0397	-0.0871	-0.191***
	(0.138)	(0.0580)	(0.0810)	(0.121)	(0.0590)	(0.0693)
Muslim OBC	0.00121	0.00878	-0.0389	0.0568	-0.115*	-0.145*
	(0.102)	(0.0624)	(0.0810)	(0.0957)	(0.0623)	(0.0801)
Other	-0.0566	0.0843*	0.0719	-0.00571	-0.0792	-0.149
	(0.0735)	(0.0431)	(0.0696)	(0.118)	(0.0762)	(0.126)
Per Capita Consumption (2005)	3.57e-06	-5.54e-06	-1.20e-05	-1.38e-05	1.57e-05	-3.01e-05
	(5.19e-05)	(3.02e-05)	(3.45e-05)	(5.11e-05)	(1.78e-05)	(3.45e-05)
Poor Household (2005)	0.0894	-0.109***	-0.0812**	-0.156**	-0.0344	-0.0428
	(0.0909)	(0.0422)	(0.0363)	(0.0762)	(0.0363)	(0.0346)
North-East	0.102	-0.141	-0.151**	-0.246	-0.111*	-0.00128
	(0.163)	(0.0866)	(0.0742)	(0.337)	(0.0583)	(0.0688)
Central	-0.0936	-0.186***	-0.105**	-0.239**	-0.200***	-0.120***
	(0.118)	(0.0445)	(0.0429)	(0.0966)	(0.0355)	(0.0350)
Western	-0.164**	-0.162***	-0.110**	-0.217***	-0.285***	-0.166***
	(0.0760)	(0.0498)	(0.0509)	(0.0787)	(0.0410)	(0.0438)
Eastern	-0.106	-0.0689	-0.0470	-0.0672	-0.161***	-0.119***
	(0.0869)	(0.0461)	(0.0474)	(0.0743)	(0.0363)	(0.0407)
Southern	-0.188**	-0.124**	-0.119***	-0.216***	-0.286***	-0.197***
	(0.0809)	(0.0500)	(0.0412)	(0.0725)	(0.0540)	(0.0398)
URBAN	0.0217	0.168***	0.0502	0.125**	0.0513*	0.0847***
	(0.0497)	(0.0336)	(0.0372)	(0.0513)	(0.0283)	(0.0296)
Constant	1.076**	0.564**	0.580**	0.204	1.204***	0.270
	(0.434)	(0.226)	(0.238)	(0.386)	(0.197)	(0.208)

	Treatment models (reference group: Not allowed to work)			
	General score Girls		**General score Boys**	
	Allowed to work and not working	**Working**	**Allowed to work and not working**	**Working**
Mother’s age	-0.0182*	0.0173	-0.0232**	-0.00427
	(0.0110)	(0.0111)	(0.0105)	(0.0108)
Hindu OBC	0.0681	0.438**	0.104	0.623***
	(0.179)	(0.192)	(0.164)	(0.183)
SCST	0.0587	0.896***	0.277	1.372***
	(0.194)	(0.206)	(0.183)	(0.197)
Muslim Upper Caste	-0.373	-1.208***	-0.488**	-1.075***
	(0.241)	(0.304)	(0.229)	(0.298)
Muslim OBC	-0.828***	-0.941***	-0.660***	-0.504**
	(0.237)	(0.271)	(0.229)	(0.255)
Other	0.708	0.0500	0.0916	-1.196*
	(0.442)	(0.491)	(0.475)	(0.656)
Per Capita Consumption (2005)	-0.000121	-0.000708***	-0.000114	-0.000319**
	(0.000146)	(0.000199)	(8.82e-05)	(0.000126)
Poor Household (2005)	0.0199	0.139	-0.0733	0.181
	(0.155)	(0.163)	(0.149)	(0.154)
Educ_GM1	0.0256	-0.00647	0.0499**	0.0666***
	(0.0228)	(0.0270)	(0.0230)	(0.0258)
Educ_GM2	-0.0130	0.0196	-0.0288	-0.0238
	(0.0267)	(0.0308)	(0.0261)	(0.0287)
Educ_male	-0.0267*	-0.0724***	-0.0204	-0.0910***
	(0.0143)	(0.0145)	(0.0142)	(0.0148)
Coverhead	-0.167	-0.559**	-0.299	-0.579***
	(0.215)	(0.221)	(0.204)	(0.213)
Land_dec	0.457	1.937***	0.0626	1.484**
	(0.577)	(0.628)	(0.605)	(0.646)
Health_dec	0.535*	-0.612**	0.554**	-0.407
	(0.281)	(0.294)	(0.279)	(0.290)
Husb_violence1	-0.0873	-0.0797	0.121	0.162
	(0.242)	(0.249)	(0.234)	(0.245)
Husb_violence2	-0.0131	-0.357	-0.300	-0.724***
	(0.242)	(0.244)	(0.232)	(0.240)
North-East	0.787	0.555	-0.0710	-0.928*
	(0.658)	(0.745)	(0.455)	(0.506)
Central	0.615***	0.594***	0.491**	0.714***
	(0.206)	(0.216)	(0.194)	(0.203)
Western	-1.195***	-1.007***	-1.086***	-0.646***
	(0.209)	(0.224)	(0.193)	(0.205)
Eastern	-0.861***	-1.209***	-0.669***	-1.044***

	(0.189)	(0.206)	(0.176)	(0.195)
Southern	-1.534***	-0.428*	-1.686***	-0.479**
	(0.259)	(0.260)	(0.238)	(0.239)
URBAN	-0.391***	-1.135***	-0.750***	-1.252***
	(0.135)	(0.145)	(0.128)	(0.138)
Constant	1.889**	1.074	2.558***	1.658**
	(0.744)	(0.782)	(0.741)	(0.794)
Observations	2,821	2,821	3,203	3,203

Source: Author's calculations from IHDS

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Tables B. Mathematics score (outcome and treatment models)

	Mathematics score girls			Mathematics score boys		
	Outcome models					
	Not allowed to work and not working	Allowed to work and not working	Working	Not allowed to work and not working	Allowed to work and not working	Working
Average treatment effect	N.a.	-0.135* (0.0739)	-0.152** (0.0730)	N.a.	0.0898 (0.0560)	0.0700 (0.0552)
Age	0.0743**	0.0938***	0.0632***	0.116***	0.0582***	0.128***
	(0.0359)	(0.0199)	(0.0210)	(0.0321)	(0.0179)	(0.0172)
Education kindergarden	0.388**	0.584***	0.624***	0.330*	0.309**	0.595***
	(0.177)	(0.138)	(0.158)	(0.176)	(0.120)	(0.110)
Education primary	0.518***	0.740***	0.812***	0.331*	0.457***	0.694***
	(0.187)	(0.143)	(0.164)	(0.193)	(0.128)	(0.114)
Education lower secondary	0.464**	0.652***	0.706***	0.380*	0.511***	0.666***
	(0.220)	(0.154)	(0.171)	(0.207)	(0.132)	(0.120)
Mother's age	-0.00611	-0.0115***	-0.00919**	-0.00531	-0.00483	-0.00394
	(0.00597)	(0.00378)	(0.00368)	(0.00591)	(0.00308)	(0.00302)
School distance	0.00204	0.0154**	0.00294	0.00783	0.00539	0.00671***
	(0.00826)	(0.00690)	(0.00550)	(0.00612)	(0.00401)	(0.00227)
Number of sisters	-0.0920**	-0.0268	-0.0989***	-0.0495	-0.0799***	-0.0195
	(0.0452)	(0.0181)	(0.0202)	(0.0380)	(0.0189)	(0.0204)
Number of brothers	-0.0504	-0.0218	-0.0856***	-0.103**	-0.0626***	-0.0149
	(0.0447)	(0.0244)	(0.0257)	(0.0449)	(0.0221)	(0.0204)
Number of days ill	-0.00332	0.000623	0.0109	-0.0241*	0.00170	-0.0119
	(0.0116)	(0.00599)	(0.00697)	(0.0136)	(0.00492)	(0.00789)
Household head full-time work	-0.181**	0.111	0.174*	-0.0610	0.00126	0.0798
	(0.0898)	(0.0889)	(0.0945)	(0.119)	(0.0545)	(0.0734)
Household head part-time work	-0.0493	0.0757	0.244***	-0.0189	0.0381	0.109
	(0.0771)	(0.0845)	(0.0933)	(0.108)	(0.0505)	(0.0715)
Highest Female Education in household	0.0102	0.0193***	0.0244***	0.0178**	0.0188***	0.0230***
	(0.00851)	(0.00548)	(0.00597)	(0.00784)	(0.00434)	(0.00412)
Highest Male Education in Household	0.0184**	0.0194***	0.0143***	0.00836	0.0125***	0.0215***
	(0.00801)	(0.00522)	(0.00536)	(0.00793)	(0.00443)	(0.00438)
Hindu OBC	0.0490	-0.0275	0.0193	0.0359	0.0127	-0.0322
	(0.0769)	(0.0488)	(0.0563)	(0.0728)	(0.0405)	(0.0444)

SCST	-0.0761	-0.115**	-0.0837	-0.215**	-0.0431	-0.123***
	(0.0845)	(0.0552)	(0.0608)	(0.100)	(0.0424)	(0.0445)
Muslim Upper Caste	-0.293*	-0.0333	0.189	-0.168	-0.0775	-0.250***
	(0.175)	(0.0750)	(0.116)	(0.117)	(0.0738)	(0.0755)
Muslim OBC	-0.130	-0.0373	-0.0175	-0.0212	-0.107	-0.114
	(0.117)	(0.0759)	(0.0993)	(0.103)	(0.0711)	(0.0948)
Other	0.0710	0.0180	0.0659	0.0474	-0.0281	-0.142
	(0.0955)	(0.0631)	(0.0871)	(0.152)	(0.0843)	(0.0998)
Per Capita Consumption (2005)	5.73e-05	1.25e-05	3.14e-05	-1.26e-05	1.38e-05	-2.82e-05
	(5.94e-05)	(3.93e-05)	(4.39e-05)	(2.89e-05)	(2.06e-05)	(3.16e-05)
Poor2005	0.0941	-0.135***	-0.0103	-0.106	-0.0300	-0.0746*
	(0.104)	(0.0524)	(0.0478)	(0.0844)	(0.0419)	(0.0410)
North-East	0.421	-0.138	-0.0729	-0.323	-0.189**	-0.0188
	(0.287)	(0.142)	(0.0962)	(0.443)	(0.0833)	(0.0853)
Central	-0.110	-0.267***	-0.262***	-0.350***	-0.252***	-0.178***
	(0.142)	(0.0534)	(0.0576)	(0.106)	(0.0436)	(0.0442)
Western	-0.203**	-0.283***	-0.207***	-0.339***	-0.422***	-0.302***
	(0.0901)	(0.0674)	(0.0620)	(0.0976)	(0.0551)	(0.0549)
Eastern	-0.0359	-0.0547	-0.0709	-0.177**	-0.147***	-0.102**
	(0.0977)	(0.0524)	(0.0637)	(0.0801)	(0.0450)	(0.0495)
Southern	-0.171*	-0.101	-0.0739	-0.267***	-0.270***	-0.166***
	(0.0898)	(0.0681)	(0.0514)	(0.0852)	(0.0600)	(0.0456)
URBAN	0.0717	0.186***	0.0399	0.151**	0.0576*	0.0593*
	(0.0646)	(0.0403)	(0.0508)	(0.0608)	(0.0344)	(0.0358)
Constant						

	Treatment models (reference group: Not allowed to work)			
	Mathematics Score Girls		**Mathematics Score Boys**	
	Allowed to work and not working	**Working**	**Allowed to work and not working**	**Working**
Mother's age	-0.0165	0.0184*	-0.0225**	-0.00293
	(0.0109)	(0.0110)	(0.0104)	(0.0107)
Hindu OBC	0.0644	0.450**	0.118	0.641***
	(0.178)	(0.191)	(0.162)	(0.181)
SCST	0.0618	0.909***	0.301*	1.396***
	(0.192)	(0.205)	(0.181)	(0.196)
Muslim Upper Caste	-0.391	-1.229***	-0.461**	-1.065***
	(0.239)	(0.303)	(0.227)	(0.298)
Muslim OBC	-0.786***	-0.849***	-0.668***	-0.514**
	(0.237)	(0.269)	(0.226)	(0.253)
Other	0.691	0.0259	0.170	-0.967
	(0.434)	(0.487)	(0.464)	(0.625)
Per Capita Consumption (2005)	-0.000111	-0.000664***	-0.000153**	-0.000338***
	(0.000146)	(0.000195)	(6.65e-05)	(0.000117)
Poor2005	0.0209	0.173	-0.0818	0.179
	(0.154)	(0.161)	(0.146)	(0.152)
North-East	0.763	0.563	-0.0618	-0.892*
	(0.655)	(0.741)	(0.459)	(0.512)
Central	0.632***	0.602***	0.439**	0.659***
	(0.205)	(0.214)	(0.191)	(0.201)
Western	-1.204***	-1.012***	-1.072***	-0.597***
	(0.208)	(0.223)	(0.194)	(0.205)
Eastern	-0.872***	-1.203***	-0.677***	-1.050***
	(0.188)	(0.204)	(0.176)	(0.196)
Southern	-1.556***	-0.400	-1.617***	-0.410*
	(0.256)	(0.256)	(0.237)	(0.239)
Educ_GM1	0.0269	-0.00773	0.0497**	0.0625**
	(0.0226)	(0.0266)	(0.0226)	(0.0252)
Educ_GM2	-0.00925	0.0158	-0.0320	-0.0298
	(0.0265)	(0.0306)	(0.0257)	(0.0284)
Educ_male	-0.0271*	-0.0701***	-0.0206	-0.0917***
	(0.0143)	(0.0144)	(0.0142)	(0.0147)
Coverhead	-0.199	-0.534**	-0.254	-0.520**
	(0.213)	(0.218)	(0.204)	(0.212)
Land_dec	0.519	2.016***	-0.0570	1.433**
	(0.573)	(0.624)	(0.601)	(0.640)
Health_dec	0.494*	-0.631**	0.573**	-0.361
	(0.280)	(0.294)	(0.276)	(0.288)

Husb_violence1	-0.124	-0.116	0.0856	0.151
	(0.240)	(0.246)	(0.233)	(0.244)
Husb_violence2	0.0781	-0.309	-0.311	-0.760***
	(0.239)	(0.242)	(0.230)	(0.237)
URBAN	-0.392***	-1.120***	-0.748***	-1.242***
	(0.134)	(0.143)	(0.127)	(0.137)
Constant	1.806**	0.894	2.644***	1.608**
	(0.739)	(0.779)	(0.732)	(0.787)
Observations	2,857	2,857	3,230	3,230

Source: Author's calculations from IHDS
Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Tables C. School hours (outcome and treatment models)

	School_hours_girls			School_hours_boys		
	Outcome models					
	Not allowed to work and not working	**Allowed to work and not working**	**Working**	**Not allowed to work and not working**	**Allowed to work and not working**	**Working**
Average treatment effect	N.a.	0.0104 (0.363)	0.368 (0.365)	N.a.	-0.514 (0.371)	0.0235 (0.376)
Age	-0.00656	0.00538	0.00338	0.00847	0.00183	-0.00325
	(0.00727)	(0.00339)	(0.00383)	(0.00657)	(0.00341)	(0.00527)
Education kindergarden	0.0861	0.0110	0.0735**	-0.0151	0.0259	0.00830
	(0.0693)	(0.0213)	(0.0286)	(0.0458)	(0.0208)	(0.0234)
Education primary	0.142*	-0.0227	0.0769**	-0.00887	0.0491**	0.0436
	(0.0767)	(0.0272)	(0.0333)	(0.0570)	(0.0248)	(0.0315)
Education lower secondary	0.177**	0.00349	0.0775**	-0.0133	0.0244	0.0553
	(0.0809)	(0.0265)	(0.0348)	(0.0597)	(0.0262)	(0.0349)
Mother’s age	0.198**	0.127*	-0.00151	-0.0897	-0.00850	0.0905
	(0.101)	(0.0712)	(0.104)	(0.210)	(0.0800)	(0.0754)
School distance	0.000969	0.00224**	0.000391	-0.00418**	0.00333***	0.000210
	(0.00187)	(0.000974)	(0.00103)	(0.00186)	(0.000858)	(0.00107)
Number of sisters	-0.00397	-0.00740**	-0.00681***	0.000805	-0.00165	-0.00471**
	(0.00386)	(0.00357)	(0.00223)	(0.00263)	(0.00170)	(0.00234)
Number of brothers	-0.00621	0.00196	0.00584	0.00285	0.00594	0.00597
	(0.0107)	(0.00451)	(0.00414)	(0.00997)	(0.00473)	(0.00633)
Number of days ill	0.0142	0.00478	0.00695	0.0146	0.00291	0.00950
	(0.0132)	(0.00585)	(0.00630)	(0.00933)	(0.00482)	(0.00585)
Household head full-time work	-0.00912*	-0.00396*	0.00213	0.00301	-0.00288	-0.000975
	(0.00527)	(0.00213)	(0.00167)	(0.00303)	(0.00259)	(0.00200)
Household head part-time work	0.00176	-0.0361**	0.0176	4.78e-05	0.00425	0.0236
	(0.0308)	(0.0174)	(0.0241)	(0.0348)	(0.0181)	(0.0305)
Highest Female Education in household	0.00603	-0.0738***	-0.00844	-0.00308	-0.0270	0.0280
	(0.0293)	(0.0165)	(0.0241)	(0.0331)	(0.0176)	(0.0303)
Highest Male Education in Household	0.000352	0.00239	0.00398**	-0.000445	0.00134	0.00441**
	(0.00228)	(0.00150)	(0.00158)	(0.00193)	(0.00148)	(0.00181)
Hindu OBC	-0.000311	0.00162	0.000800	0.00295	0.00169	0.00209
	(0.00238)	(0.00156)	(0.00139)	(0.00221)	(0.00125)	(0.00168)

SCST	0.0109	0.00460	0.0367	0.0158	0.0293**	0.0463**
	(0.0242)	(0.0147)	(0.0223)	(0.0247)	(0.0143)	(0.0201)
Muslim Upper Caste	-0.0117	0.0260*	0.0254	-0.00705	0.0242	0.0176
	(0.0263)	(0.0153)	(0.0216)	(0.0271)	(0.0149)	(0.0195)
Muslim OBC	-0.0777**	-0.0757***	-0.0660*	-0.0632**	-0.0466*	-0.0120
	(0.0368)	(0.0241)	(0.0391)	(0.0320)	(0.0247)	(0.0464)
Other	-0.0109	-0.0285	-0.00504	-0.0334	0.0335*	0.0274
	(0.0342)	(0.0241)	(0.0309)	(0.0291)	(0.0196)	(0.0262)
Per Capita Consumption (2005)	0.0565*	-0.0119	0.0127	0.0208	-0.00315	-0.000192
	(0.0314)	(0.0222)	(0.0354)	(0.0301)	(0.0253)	(0.0492)
Poor2005	-2.98e-05	2.61e-05**	3.63e-06	-1.15e-05	1.34e-05	4.29e-06
	(2.87e-05)	(1.08e-05)	(1.40e-05)	(1.75e-05)	(1.18e-05)	(7.46e-06)
North-East	-0.0494*	-0.0252*	0.00609	-0.0449*	-0.00824	-0.000269
	(0.0267)	(0.0135)	(0.0118)	(0.0236)	(0.0120)	(0.0113)
Central	0.0796	-0.230***	-0.307***	0.0443	-0.304***	-0.229***
	(0.0768)	(0.0387)	(0.0921)	(0.0689)	(0.0427)	(0.0486)
Western	-0.0156	0.00542	-0.0134	0.0435*	-0.0184*	-0.0127
	(0.0294)	(0.00972)	(0.0130)	(0.0264)	(0.00944)	(0.0138)
Eastern	-0.0228	-0.104***	-0.104***	-0.0456	-0.0473***	-0.0632**
	(0.0268)	(0.0210)	(0.0289)	(0.0289)	(0.0160)	(0.0255)
Southern	-0.137***	-0.150***	-0.181***	-0.0599**	-0.141***	-0.185***
	(0.0287)	(0.0137)	(0.0183)	(0.0267)	(0.0140)	(0.0188)
URBAN	-0.0397	-0.0615***	0.0152	0.0200	-0.0298	0.00515
	(0.0310)	(0.0234)	(0.0161)	(0.0332)	(0.0210)	(0.0149)
Constant	3.479***	3.414***	3.363***	3.537***	3.326***	3.441***
	(0.105)	(0.0542)	(0.0627)	(0.0950)	(0.0485)	(0.0609)

	Treatment models (reference group: not allowed to work)			
	School_hours_girls		School_hours_boys	
	Allowed to work and not working	Working	Allowed to work and not working	Working
Mother's age	-0.0200***	0.00556	-0.0216***	-0.00781
	(0.00686)	(0.00684)	(0.00615)	(0.00627)
Hindu OBC	0.153	0.443***	0.148	0.732***
	(0.111)	(0.119)	(0.102)	(0.114)
SCST	0.133	0.944***	0.334***	1.374***
	(0.122)	(0.128)	(0.114)	(0.123)
Muslim Upper Caste	-0.204	-0.865***	-0.314**	-0.764***
	(0.152)	(0.185)	(0.144)	(0.183)
Muslim OBC	-0.636***	-0.704***	-0.638***	-0.528***
	(0.152)	(0.169)	(0.144)	(0.161)
Other	0.276	-0.251	0.516*	0.0873
	(0.258)	(0.297)	(0.268)	(0.314)
Per Capita Consumption (2005)	-0.000165**	-0.000532***	-0.000111**	-0.000232***
	(7.79e-05)	(0.000109)	(4.37e-05)	(8.78e-05)
Poor2005	-0.0486	0.184*	-0.0979	0.213**
	(0.0974)	(0.101)	(0.0911)	(0.0964)
North-East	-0.507**	-1.224***	-0.616***	-1.271***
	(0.244)	(0.298)	(0.220)	(0.274)
Central	0.594***	0.652***	0.378***	0.536***
	(0.136)	(0.142)	(0.124)	(0.131)
Western	-1.085***	-0.714***	-1.094***	-0.678***
	(0.135)	(0.140)	(0.125)	(0.133)
Eastern	-0.768***	-1.206***	-0.751***	-1.016***
	(0.117)	(0.129)	(0.113)	(0.125)
Southern	-1.481***	-0.286*	-1.507***	-0.256*
	(0.156)	(0.157)	(0.150)	(0.151)
Educ_GM1	0.0295**	-0.0146	0.0424***	0.0383**
	(0.0146)	(0.0172)	(0.0149)	(0.0168)
Educ_GM2	0.0100	0.0367*	-0.0256	-0.0464**
	(0.0177)	(0.0208)	(0.0162)	(0.0193)
Educ_male	-0.0252***	-0.0819***	-0.00953	-0.0882***
	(0.00887)	(0.00916)	(0.00868)	(0.00913)
Coverhead	0.0319	-0.337**	-0.0709	-0.477***
	(0.134)	(0.137)	(0.127)	(0.132)
Land_dec	0.848**	1.747***	0.475	1.616***
	(0.366)	(0.400)	(0.382)	(0.396)
Health_dec	0.300*	-0.723***	0.549***	-0.577***
	(0.181)	(0.189)	(0.175)	(0.182)

Husb_violence1	0.0661	0.113	0.353**	0.439***
	(0.156)	(0.160)	(0.150)	(0.156)
Husb_violence2	0.0305	-0.352**	-0.373**	-0.809***
	(0.158)	(0.160)	(0.149)	(0.154)
URBAN	-0.537***	-1.175***	-0.705***	-1.248***
	(0.0829)	(0.0890)	(0.0785)	(0.0851)
Constant	1.588***	1.400***	1.818***	1.536***
	(0.459)	(0.487)	(0.442)	(0.468)
Observations	7,074	7,074	7,861	7,861

Source: Author’s calculations from IHDS
Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Tables D. Homework hours (outcome and treatment models)

	Homework hours girls			Homework hours boys		
	Outcome models					
	Not allowed to work and not working	**Allowed to work and not working**	**Working**	**Not allowed to work and not working**	**Allowed to work and not working**	**Working**
Average treatment effect	N.a.	-0.0355 (0.283)	-0.435 (0.272)	N.a.	-0.311 (0.260)	-0.460* (0.256)
Age	0.0336*	0.0103	0.00794	0.0348*	0.00404	0.00971
	(0.0198)	(0.0102)	(0.0108)	(0.0188)	(0.00955)	(0.0110)
Education kindergarden	0.171	0.128	0.238***	0.187	0.158***	0.0312
	(0.219)	(0.0789)	(0.0756)	(0.117)	(0.0514)	(0.0810)
Education primary	0.358	0.203**	0.352***	0.164	0.277***	0.0944
	(0.224)	(0.0885)	(0.0884)	(0.142)	(0.0644)	(0.0921)
Education lower secondary	0.287	0.298***	0.405***	0.272*	0.349***	0.183*
	(0.235)	(0.0924)	(0.0916)	(0.149)	(0.0729)	(0.0987)
Mother’s age	0.0643	0.406***	0.752***	-0.225	0.125	0.688**
	(0.405)	(0.155)	(0.121)	(0.266)	(0.181)	(0.285)
School distance	-0.00412	-0.000702	0.00264	-0.00589	-0.00312	0.00615**
	(0.00535)	(0.00267)	(0.00305)	(0.00444)	(0.00256)	(0.00252)
Number of sisters	0.0262***	0.00179	-0.000421	0.00877	0.0111**	0.00242
	(0.00802)	(0.00611)	(0.00325)	(0.00540)	(0.00458)	(0.00334)
Number of brothers	-0.0933***	-0.0430***	-0.0376**	-0.0568	-0.0364**	-0.0172
	(0.0328)	(0.0152)	(0.0149)	(0.0347)	(0.0158)	(0.0202)
Number of days ill	-0.00134	-0.0342*	-0.0497***	-0.0632*	-0.0481***	-0.0561***
	(0.0329)	(0.0205)	(0.0183)	(0.0357)	(0.0180)	(0.0161)
Household head full-time work	-0.00926	0.00167	-0.00249	-0.00313	-0.00489	-0.00419
	(0.00770)	(0.00479)	(0.00573)	(0.00619)	(0.00480)	(0.00501)
Household head part-time work	-0.0424	0.0358	-0.00266	-0.0673	-0.0711	-0.0674
	(0.0746)	(0.0574)	(0.0621)	(0.0747)	(0.0470)	(0.0611)
Highest Female Education in household	-0.0165	0.00882	0.00708	-0.0983	-0.0694	-0.0312
	(0.0677)	(0.0544)	(0.0613)	(0.0697)	(0.0443)	(0.0619)
Highest Male Education in Household	-0.000132	0.00935**	0.0141***	0.00847	0.0139***	0.0244***
	(0.00591)	(0.00421)	(0.00487)	(0.00713)	(0.00416)	(0.00409)
Hindu OBC	0.00900	0.0187***	0.0125***	0.0129*	0.0101**	0.0116***
	(0.00691)	(0.00456)	(0.00407)	(0.00685)	(0.00396)	(0.00403)

SCST	-0.0131	-0.0115	-0.165***	0.0882*	0.0403	-0.0961**
	(0.0569)	(0.0401)	(0.0581)	(0.0526)	(0.0403)	(0.0454)
Muslim Upper Caste	-0.188***	-0.0706	-0.105*	-0.0254	-0.0264	-0.0786*
	(0.0725)	(0.0537)	(0.0581)	(0.0791)	(0.0419)	(0.0475)
Muslim OBC	-0.147	0.0318	-0.144	0.0822	0.123**	-0.0746
	(0.0988)	(0.0602)	(0.103)	(0.0751)	(0.0569)	(0.0877)
Other	-0.154*	-0.157*	-0.358***	0.0741	0.00905	-0.293***
	(0.0872)	(0.0899)	(0.0942)	(0.0834)	(0.0656)	(0.0838)
Per Capita Consumption (2005)	-0.262**	-0.00229	0.113	0.138	0.0181	-0.140
	(0.120)	(0.103)	(0.180)	(0.185)	(0.0781)	(0.0890)
Poor2005	1.71e-05	5.64e-05*	-8.00e-05*	-1.76e-05	-2.78e-06	3.10e-05*
	(5.11e-05)	(3.25e-05)	(4.72e-05)	(3.06e-05)	(2.82e-05)	(1.59e-05)
North-East	-0.0496	-0.00957	-0.0738*	-0.163**	-0.0370	0.0303
	(0.0702)	(0.0485)	(0.0378)	(0.0669)	(0.0368)	(0.0309)
Central	0.542***	0.0891	0.238*	0.450***	0.0203	0.103
	(0.184)	(0.0663)	(0.136)	(0.173)	(0.0697)	(0.115)
Western	-0.0679	-0.0213	0.00345	0.0535	-0.0112	0.0228
	(0.0847)	(0.0344)	(0.0383)	(0.0723)	(0.0318)	(0.0349)
Eastern	0.327***	0.150*	0.143**	0.0852	0.186***	0.195***
	(0.0944)	(0.0792)	(0.0599)	(0.0712)	(0.0502)	(0.0540)
Southern	0.346***	0.179***	0.206***	0.285***	0.141***	0.159***
	(0.0714)	(0.0367)	(0.0599)	(0.0759)	(0.0338)	(0.0525)
URBAN	0.188**	0.0788	0.0896*	0.0774	0.0442	0.202***
	(0.0857)	(0.0641)	(0.0477)	(0.101)	(0.0649)	(0.0428)
Constant	1.682***	1.664***	1.626***	1.753***	1.882***	1.556***
	(0.306)	(0.148)	(0.165)	(0.264)	(0.138)	(0.161)

	Treatment models (reference group: not allowed to work)			
	Homework hours girls		Homework hours boys	
	Allowed to work and not working	Working	Allowed to work and not working	Working
Mother's age	-0.0205***	0.00505	-0.0227***	-0.00887
	(0.00688)	(0.00686)	(0.00615)	(0.00628)
Hindu OBC	0.147	0.438***	0.135	0.727***
	(0.111)	(0.120)	(0.102)	(0.114)
SCST	0.132	0.943***	0.330***	1.384***
	(0.122)	(0.129)	(0.114)	(0.124)
Muslim Upper Caste	-0.220	-0.867***	-0.343**	-0.742***
	(0.153)	(0.186)	(0.144)	(0.183)
Muslim OBC	-0.629***	-0.682***	-0.658***	-0.535***
	(0.153)	(0.169)	(0.144)	(0.161)
Other	0.235	-0.284	0.534**	0.114
	(0.255)	(0.294)	(0.268)	(0.314)
Per Capita Consumption (2005)	-0.000163**	-0.000537***	-0.000107**	-0.000233***
	(7.76e-05)	(0.000110)	(4.37e-05)	(8.86e-05)
Poor2005	-0.0576	0.183*	-0.0915	0.220**
	(0.0977)	(0.102)	(0.0914)	(0.0967)
North-East	-0.495**	-1.236***	-0.605***	-1.269***
	(0.244)	(0.298)	(0.220)	(0.275)
Central	0.585***	0.646***	0.371***	0.537***
	(0.136)	(0.142)	(0.124)	(0.131)
Western	-1.109***	-0.752***	-1.107***	-0.687***
	(0.134)	(0.140)	(0.125)	(0.133)
Eastern	-0.764***	-1.213***	-0.738***	-1.007***
	(0.117)	(0.129)	(0.113)	(0.125)
Southern	-1.462***	-0.291*	-1.513***	-0.258*
	(0.157)	(0.158)	(0.151)	(0.152)
Educ_GM1	0.0276*	-0.0162	0.0408***	0.0362**
	(0.0145)	(0.0172)	(0.0148)	(0.0167)
Educ_GM2	0.00878	0.0345*	-0.0282*	-0.0476**
	(0.0176)	(0.0208)	(0.0161)	(0.0192)
Educ_male	-0.0249***	-0.0818***	-0.0100	-0.0878***
	(0.00893)	(0.00924)	(0.00870)	(0.00916)
Coverhead	0.0285	-0.352**	-0.0583	-0.463***
	(0.135)	(0.138)	(0.127)	(0.132)
Land_dec	0.964***	1.810***	0.450	1.582***
	(0.366)	(0.400)	(0.383)	(0.396)
Health_dec	0.286	-0.739***	0.547***	-0.574***
	(0.181)	(0.189)	(0.175)	(0.182)

Husb_violence1	0.0628	0.106	0.337**	0.429***
	(0.156)	(0.160)	(0.150)	(0.156)
Husb_violence2	0.0495	-0.339**	-0.372**	-0.810***
	(0.158)	(0.160)	(0.149)	(0.154)
URBAN	-0.537***	-1.185***	-0.697***	-1.241***
	(0.0831)	(0.0895)	(0.0786)	(0.0852)
Constant	1.513***	1.402***	1.889***	1.591***
	(0.461)	(0.490)	(0.443)	(0.469)
Observations	7,016	7,016	7,816	7,816

Source: Author’s calculations from IHDS
Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Tables E. Days of absence (outcome and treatment models)

	Days absence girls			Days absence boys		
	Outcome models					
	Not allowed to work and not working	**Allowed to work and not working**	**Working**	**Not allowed to work and not working**	**Allowed to work and not working**	**Working**
Average treatment effect	N.a.	-0.0400 (0.224)	-0.109 (0.219)	N.a.	0.0104 (0.213)	0.264 (0.213)
Age	0.00453	0.0251	-0.000863	0.0154	-0.00825	0.0170
	(0.0383)	(0.0178)	(0.0208)	(0.0377)	(0.0169)	(0.0163)
Education kindergarden	0.0981	-0.227**	-0.382***	-0.423**	-0.0872	-0.134
	(0.248)	(0.0935)	(0.103)	(0.181)	(0.0953)	(0.0912)
Education primary	0.243	-0.343***	-0.383***	-0.106	-0.204*	-0.183
	(0.341)	(0.133)	(0.137)	(0.250)	(0.118)	(0.143)
Education lower secondary	-0.155	-0.410***	-0.452***	-0.389	-0.0934	-0.387***
	(0.345)	(0.128)	(0.148)	(0.254)	(0.132)	(0.127)
Mother's age	0.546	-0.981*	0.520	-0.716	0.763**	-0.0140
	(0.488)	(0.512)	(0.375)	(0.476)	(0.383)	(0.674)
School distance	0.00193	-0.00457	-0.000987	0.00362	-0.000842	0.00880*
	(0.00731)	(0.00569)	(0.00691)	(0.00983)	(0.00551)	(0.00477)
Number of sisters	-0.00191	0.0121	0.00684	0.00575	0.0131	0.00253
	(0.0167)	(0.00967)	(0.00656)	(0.0125)	(0.00805)	(0.00744)
Number of brothers	-0.0401	-0.0335	0.0686**	-0.00287	0.00990	-0.00522
	(0.0491)	(0.0271)	(0.0278)	(0.0529)	(0.0229)	(0.0330)
Number of days ill	0.0329	0.000389	0.00106	-0.0421	-0.0136	-0.00500
	(0.0531)	(0.0276)	(0.0290)	(0.0575)	(0.0267)	(0.0322)
Household head full-time work	0.0808***	0.0439***	0.0499***	0.0233	0.0388***	0.0354***
	(0.0111)	(0.00773)	(0.00690)	(0.0149)	(0.00661)	(0.00613)
Household head part-time work	0.0784	0.0891	-0.0875	-0.549***	-0.0412	-0.257*
	(0.152)	(0.102)	(0.101)	(0.189)	(0.0963)	(0.137)
Highest Female Education in household	-0.0285	-0.0143	-0.0456	-0.283*	-0.124	-0.326**
	(0.141)	(0.100)	(0.107)	(0.172)	(0.0955)	(0.136)
Highest Male Education in Household	-0.0448***	-0.0380***	-0.0127	-0.0268**	-0.0283***	-0.0247**
	(0.0122)	(0.00813)	(0.00864)	(0.0131)	(0.00807)	(0.0104)
Hindu OBC	-0.00308	-0.00671	-0.0147**	-0.0251*	0.00562	-0.0132
	(0.0124)	(0.00659)	(0.00629)	(0.0135)	(0.00657)	(0.00843)

SCST	-0.288*	-0.0217	0.204*	-0.0331	-0.140*	0.0912
	(0.150)	(0.0925)	(0.119)	(0.154)	(0.0808)	(0.114)
Muslim Upper Caste	-0.248	-0.0217	0.148	-0.141	-0.192**	0.0763
	(0.163)	(0.0937)	(0.113)	(0.161)	(0.0875)	(0.110)
Muslim OBC	-0.161	0.00814	0.0352	0.0503	-0.0751	0.288
	(0.194)	(0.111)	(0.141)	(0.164)	(0.109)	(0.187)
Other	-0.302	0.187	0.324**	0.0225	-0.0334	0.227
	(0.222)	(0.123)	(0.146)	(0.195)	(0.117)	(0.145)
Per Capita Consumption (2005)	-1.090***	-0.290	-0.259	-0.441	0.138	0.342
	(0.388)	(0.203)	(0.276)	(0.483)	(0.253)	(0.289)
Poor2005	-0.000169	0.000160*	3.08e-05	-0.000179	2.84e-05	6.27e-05
	(0.000137)	(9.15e-05)	(0.000115)	(0.000151)	(4.76e-05)	(8.23e-05)
North-East	-0.283**	0.122*	-0.0191	-0.173	0.00565	-0.00947
	(0.131)	(0.0722)	(0.0650)	(0.131)	(0.0687)	(0.0665)
Central	0.750***	-0.100	0.117	0.0115	-0.120	0.175
	(0.180)	(0.174)	(0.200)	(0.221)	(0.138)	(0.144)
Western	0.292**	0.210***	0.0973	-0.0301	0.215***	0.161*
	(0.145)	(0.0766)	(0.0760)	(0.142)	(0.0700)	(0.0858)
Eastern	-0.271*	-0.311***	-0.322***	-0.934***	-0.397***	-0.0912
	(0.159)	(0.111)	(0.114)	(0.169)	(0.113)	(0.115)
Southern	0.152	0.300***	0.163*	-0.122	0.282***	0.212**
	(0.136)	(0.0822)	(0.0978)	(0.149)	(0.0809)	(0.0890)
URBAN	0.527***	-0.0169	0.157*	0.180	0.000244	0.139
	(0.162)	(0.138)	(0.0950)	(0.175)	(0.132)	(0.107)
Constant	1.442***	1.357***	1.463***	2.230***	1.571***	1.206***
	(0.430)	(0.274)	(0.383)	(0.525)	(0.269)	(0.322)

	Treatment models (reference group: not allowed to work)			
	Days absence girls		**Days absence boys**	
	Allowed to work and not working	**Working**	**Allowed to work and not working**	**Working**
Mother's age	-0.0187***	0.00576	-0.0205***	-0.00672
	(0.00692)	(0.00690)	(0.00624)	(0.00638)
Hindu OBC	0.152	0.438***	0.121	0.704***
	(0.112)	(0.120)	(0.103)	(0.115)
SCST	0.129	0.939***	0.339***	1.375***
	(0.122)	(0.129)	(0.115)	(0.124)
Muslim Upper Caste	-0.197	-0.883***	-0.315**	-0.810***
	(0.154)	(0.187)	(0.146)	(0.186)
Muslim OBC	-0.656***	-0.685***	-0.679***	-0.537***
	(0.157)	(0.172)	(0.146)	(0.162)
Other	0.244	-0.296	0.558**	0.0726
	(0.256)	(0.295)	(0.272)	(0.318)
Per Capita Consumption (2005)	-0.000167**	-0.000521***	-0.000137***	-0.000313***
	(7.86e-05)	(0.000109)	(4.64e-05)	(8.93e-05)
Poor2005	-0.0719	0.169*	-0.123	0.179*
	(0.0985)	(0.102)	(0.0922)	(0.0974)
North-East	-0.499**	-1.292***	-0.573***	-1.228***
	(0.244)	(0.301)	(0.222)	(0.277)
Central	0.629***	0.680***	0.406***	0.540***
	(0.137)	(0.143)	(0.125)	(0.131)
Western	-1.122***	-0.796***	-1.107***	-0.686***
	(0.135)	(0.140)	(0.127)	(0.133)
Eastern	-0.758***	-1.205***	-0.706***	-1.000***
	(0.118)	(0.130)	(0.114)	(0.125)
Southern	-1.506***	-0.262*	-1.474***	-0.210
	(0.158)	(0.159)	(0.152)	(0.154)
Educ_GM1	0.0311**	-0.0164	0.0454***	0.0414**
	(0.0147)	(0.0174)	(0.0150)	(0.0170)
Educ_GM2	0.00827	0.0366*	-0.0284*	-0.0494**
	(0.0178)	(0.0210)	(0.0163)	(0.0194)
Educ_male	-0.0268***	-0.0819***	-0.00807	-0.0850***
	(0.00896)	(0.00924)	(0.00875)	(0.00920)
Coverhead	0.0231	-0.310**	-0.0558	-0.438***
	(0.135)	(0.138)	(0.128)	(0.133)
Land_dec	0.946**	1.799***	0.505	1.629***
	(0.368)	(0.402)	(0.385)	(0.399)
Health_dec	0.292	-0.750***	0.591***	-0.543***
	(0.184)	(0.192)	(0.177)	(0.184)

Husb_violence1	0.0427	0.0442	0.376**	0.431***
	(0.158)	(0.162)	(0.152)	(0.159)
Husb_violence2	0.0557	-0.311*	-0.394***	-0.817***
	(0.159)	(0.161)	(0.151)	(0.156)
URBAN	-0.541***	-1.185***	-0.720***	-1.232***
	(0.0840)	(0.0901)	(0.0799)	(0.0859)
Constant	1.489***	1.382***	1.713***	1.483***
	(0.463)	(0.491)	(0.446)	(0.471)
Observations	6,963	6,963	7,738	7,738

Source: Author's calculations from IHDS
Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix 1.8. Balance tests

General score (girls)

	Treatment level 2				Treatment level 3			
	Standardized differences		**Variance ratio**		**Standardized differences**		**Variance ratio**	
	Raw	**Weighted**	Raw	**Weighted**	Raw	**Weighted**	Raw	Weighted
Mother's age	-0.026	**0.067**	0.964	**1.061**	0.127	**0.093**	1.110	1.030
Educ_GM1	-0.128	**-0.014**	0.792	**0.921**	-0.353	**0.107**	0.515	1.700
Educ_GM2	-0.141	**-0.014**	0.686	**0.944**	-0.308	**0.074**	0.441	1.493
Educ_male	-0.158	**-0.021**	1.001	**1.049**	-0.535	**0.014**	0.912	1.146
Hindu OBC	0.072	**0.003**	1.056	**1.002**	0.103	**-0.042**	1.078	0.970
SCST	0.140	**0.112**	1.186	**1.089**	0.572	**0.081**	1.502	1.066
Muslim Upper Caste	-0.075	**-0.059**	0.818	**0.820**	-0.348	**-0.063**	0.246	0.807
Muslim OBC	-0.201	**-0.072**	0.609	**0.803**	-0.349	**-0.062**	0.348	0.832
Other	0.054	**-0.051**	1.393	**0.729**	-0.059	**-0.042**	0.634	0.774
Coverhead	0.252	**0.033**	0.844	**0.944**	0.038	**0.034**	1.026	0.874
Land_dec	-0.005	**-0.064**	0.930	**1.174**	0.075	**-0.079**	0.832	1.204
Health_dec	0.193	**0.023**	0.806	**1.050**	0.100	**0.040**	0.803	0.987
Husb_violence1	0.060	**0.074**	0.965	**1.023**	0.145	**0.087**	0.930	1.055
Husb_violence2	-0.228	**0.000**	0.968	**1.049**	-0.153	**-0.010**	0.975	1.017
Per Capita Consumption (2005)	-0.155	**-0.085**	1.037	**1.160**	-0.587	**0.031**	0.522	1.887
Poor2005	0.105	**0.172**	1.099	**1.143**	0.344	**0.139**	1.245	1.120
North-Eastern	0.132	**-0.079**	3.268	**0.606**	0.081	**-0.058**	2.223	0.705
Central	0.535	**0.165**	2.106	**1.204**	0.532	**0.154**	2.101	1.193
Western	-0.271	**-0.025**	0.588	**0.944**	-0.286	**-0.038**	0.564	0.916
Eastern	-0.211	**-0.023**	0.760	**0.962**	-0.323	**-0.031**	0.621	0.949
Southern	-0.399	**-0.035**	0.439	**0.938**	-0.096	**-0.050**	0.868	0.913
Urban	-0.322	**-0.041**	0.926	**0.970**	-0.734	**0.017**	0.619	1.012

General score (boys)

	Treatment level 2				Treatment level 3			
	Standardized differences		**Variance ratio**		**Standardized differences**		**Variance ratio**	
	Raw	Weighted	Raw	Weighted	Raw	Weighted	Raw	Weighted
Mother's age	-0.058	**0.013**	0.993	**1.007**	0.022	**0.0725**	1.082	**1.003**
Educ_GM1	-0.119	**-0.067**	0.859	**0.830**	-0.304	**0.009**	0.603	**1.161**
Educ_GM2	-0.225	**-0.048**	0.657	**0.878**	-0.398	**0.017**	0.407	**1.181**
Educ_male	-0.151	**0.020**	0.926	**0.895**	-0.579	**0.064**	0.940	**1.029**
Hindu OBC	0.063	**0.025**	1.043	**1.017**	0.061	**0.002**	1.042	**1.001**
SCST	0.213	**0.060**	1.323	**1.043**	0.680	**0.037**	1.672	**1.028**
Muslim Upper Caste	-0.168	**-0.024**	0.623	**0.915**	-0.388	**0.003**	0.204	**1.012**
Muslim OBC	-0.207	**-0.054**	0.563	**0.839**	-0.299	**-0.069**	0.390	**0.797**
Other	-0.048	**-0.054**	0.733	**0.679**	-0.183	**0.011**	0.163	**1.075**
Coverhead	0.309	**0.053**	0.866	**0.935**	0.105	**0.061**	1.056	**0.896**
Land_dec	-0.038	**-0.050**	0.878	**0.889**	0.042	**-0.056**	0.838	**1.011**
Health_dec	0.219	**-0.017**	0.766	**1.049**	0.140	**-0.016**	0.786	**1.037**
Husb_violence1	0.077	**0.032**	1.010	**1.092**	0.149	**0.058**	0.915	**1.000**
Husb_violence2	-0.257	**-0.027**	0.864	**0.949**	-0.208	**-0.007**	0.956	**0.993**
Per Capita Consumption (2005)	-0.194	**-0.084**	0.953	**1.205**	-0.525	**-0.002**	0.525	**1.387**
Poor2005	0.040	**0.068**	1.043	**1.055**	0.333	**0.040**	1.276	**1.033**
North-Eastern	0.0534	**-0.047**	1.444	**0.722**	-0.044	**-0.066**	0.692	**0.619**
Central	0.486	**0.150**	1.942	**1.168**	0.516	**0.156**	1.982	**1.174**
Western	-0.240	**0.013**	0.606	**1.032**	-0.212	**-0.004**	0.652	**0.989**
Eastern	-0.120	**-0.048**	0.839	**0.921**	-0.269	**-0.047**	0.631	**0.924**
Southern	-0.503	**-0.070**	0.360	**0.875**	-0.197	**-0.068**	0.757	**0.878**
Urban	-0.490	-0.055	0.836	0.9547657	-0.765	-0.009	0.612	0.992

Source: Author's calculations from IHDS

Appendix to Chapter 2

Appendix 2.1. Distribution of individuals across occupational groups

Table A. Shares of men and women in occupational groups

	2005			2011-12		
	Men	**Women**	**Total**	**Men**	**Women**	**Total**
Casual	79.5	20.5	100.0	76.5	23.5	100.0
Regular	83.6	16.4	100.0	80.1	19.9	100.0
Agriculture	63.9	36.1	100.0	55.8	44.2	100.0
Manufacturing	79.2	20.8	100.0	82.9	17.1	100.0
Services	80.8	19.2	100.0	75.8	24.2	100.0
Public_Adm	86.2	13.8	100.0	90.1	9.9	100.0
Construction	88.9	11.1	100.0	83.2	16.8	100.0
Skill level 1	72.7	27.3	100.0	63.1	36.9	100.0
Skill level 2	85.4	14.6	100.0	83.6	16.4	100.0
Skill level 3	71.8	28.2	100.0	69.8	30.2	100.0
Skill level 4	91.5	8.5	100.0	91.3	8.7	100.0
Total	80.7	19.3	100.0	78.2	21.8	100.0

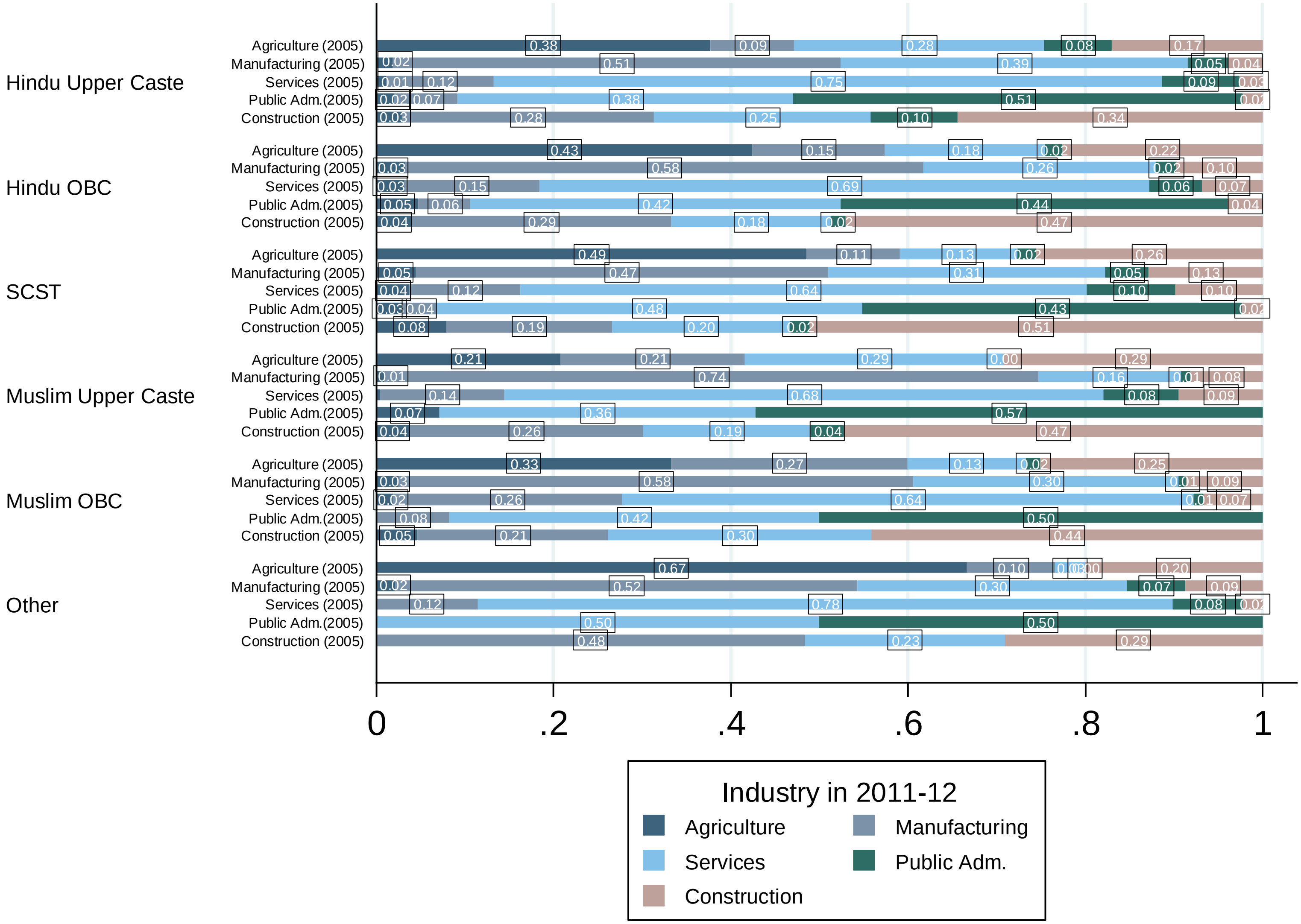
Source: Author's calculations from the IHDS dataset

Table B. Share of workers in occupational groups by religion and caste

	2005 (Row total=100%)						2011-12 (Row total=100%)					
	Upper caste Hindu	**OBC**	**SCST**	**Muslim Upper Caste**	**Muslim OBC**	**Other**	**Upper caste Hindu**	**OBC**	**SCST**	**Muslim Upper Caste**	**Muslim OBC**	**Other**
Casual	18.1	34.0	26.9	7.9	10.2	2.9	13.1	33.2	30.6	8.2	12.1	2.8
Permanent	38.6	24.9	24.3	4.0	3.8	4.4	35.1	27.6	24.2	4.7	4.8	3.5
Agriculture (2005)	8.1	32.1	43.1	4.0	9.7	2.9	8.3	36.5	41.2	3.3	8.3	2.4
Manufacturing	20.8	37.0	17.6	10.7	10.4	3.6	20.1	33.0	21.9	9.0	12.8	3.1
Services	30.3	29.8	21.6	6.5	8.0	3.9	29.3	29.5	24.5	5.7	7.4	3.6
Public_Adm	37.7	23.7	29.0	4.4	2.9	2.3	35.4	23.2	31.3	4.8	2.2	3.2
Construction	11.8	28.0	41.8	6.6	9.0	2.8	9.4	30.9	41.7	7.5	8.5	1.9
Skill level 1	18.4	32.9	33.5	4.7	8.2	2.2	15.2	31.5	36.6	5.0	9.2	2.5
Skill level 2	21.4	32.2	25.3	8.4	9.1	3.6	19.2	31.9	28.1	7.9	10.1	2.8
Skill level 3	43.5	23.0	20.0	3.7	4.9	4.9	40.8	26.1	20.0	4.3	4.2	4.6
Skill level 4	51.5	31.1	8.9	2.4	3.4	2.7	47.5	24.9	16.1	3.5	3.3	4.7
Total	24.4	31.2	26.2	6.7	8.2	3.4	23.6	30.5	27.6	6.5	8.6	3.2

Source: Author's calculations from the IHDS dataset

Appendix 2.2. Detailed Transition across industries by religion/caste between 2005 and 2011-12



Source: Author's calculations from the IHDS dataset

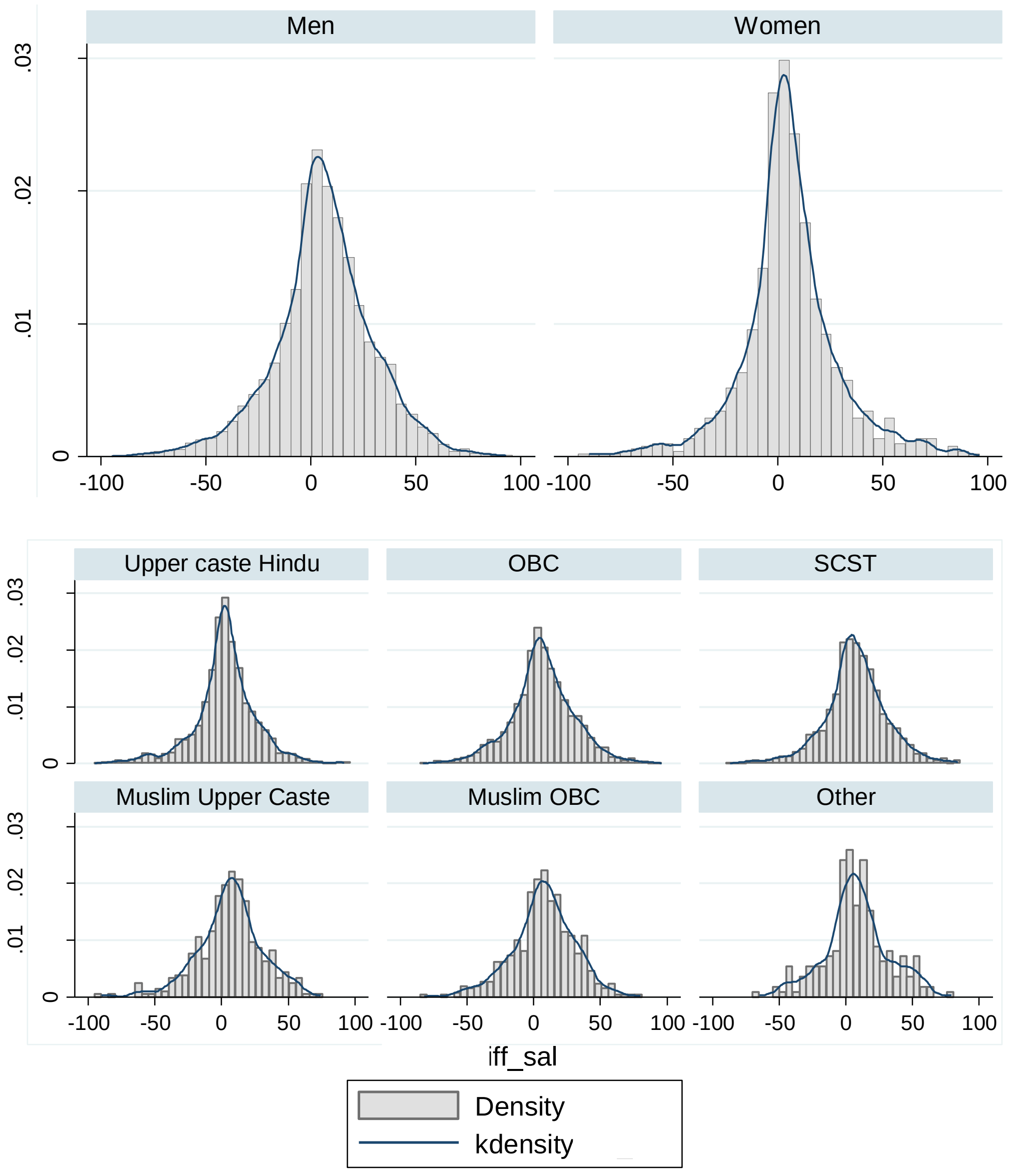
Appendix 2.3. Gender, caste and religion groups in the sample of non-missing earnings

Group	Percent (2005)	Mean hourly earnings (INR, 2005)	Percent (2011-2012)	Mean hourly earnings (INR, 2011-12)
Women	19.12	14.43	23.51	31.93
Men	80.88	21.02	76.49	41.04
Forward Caste Hindu	25.96	27.45	25.49	52.72
Other Backward Caste	31.21	16.39	31.98	35.58
Scheduled Caste and Tribes	23.74	16.76	23.67	35.17
Forward Caste Muslim	15.41[93]	14.19	6.19	30.44
OBC Muslim		12.77	9.06	26.44
Other	570	22.52	3.61	32.05

Source: Author's calculations from IHDS data

[93] The first wave of the IHDS data does not allow to distinguish between Muslim Upper Caste and Muslim OBC groups.

Appendix 2.4. Histograms of Percentile Change per group



Source: Author's calculations from the IHDS dataset
Note: The X-axis refers to Percentile Change

Appendix 2.5. Descriptive statistics for independent variables

Independent variable (from the 2005 wave)	Mean or percent	Standard deviation
Age	35.998	10.614
Educ_none	0.187	0.390
Educ_primary	0.076	0.264
Educ_middle	0.263	0.440
Educ_secondary	0.151	0.358
Educ_higher	0.244	0.430
Sec1	0.111	0.314
Sec2	0.210	0.407
Sec3	0.055	0.228
Number of children	1.587	1.463
Married	0.779	0.415
State control variable	N.a.	N.a.

Source: Author's calculations from the IHDS dataset

Appendix 2.6. Probit estimations used to generate the selection terms

Table A. Probit estimation of labor market participation

Variables (2005)	**Labor market participation**
Age	0.040***
	(0.000)
Age squared	-0.000***
	(0.000)
Educ_primary (2005)	0.018***
	(0.007)
Educ_middle (2005)	0.009**
	(0.004)
Educ_secondary (2005)	0.014
	(0.009)
Educ_higher (2005)	0.054***
	(0.011)
Sec1 (2005)	0.017
	(0.011)
Sec2 (2005)	-0.018**
	(0.009)
Sec3 (2005)	-0.019*
	(0.010)
Hindu OBC	0.062***
	(0.004)
SCST	0.118***
	(0.006)
Muslim Upper Caste	0.068***
	(0.006)
Muslim OBC	0.005
	(0.008)
Female infants	0.006**
	(0.003)
Male infants	0.005*
	(0.003)
Female child	-0.010***
	(0.002)
Male child	-0.009***
	(0.002)
Elderly male	-0.013***
	(0.004)
Elderly female	-0.040***
	(0.004)
State control variables	Yes
Observations	64,110
Pseudo-R^2	17.46
LR chi^2	12,373.79
Prob>chi^2	0.000

Source: Author's calculations from the IHDS dataset
Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B. Probit estimation of attrition

Variables	Attrition
Female (2005)	-0.110***
	(0.018)
PersonID (2005)	-0.012***
	(0.004)
Relationship to household head (2005)	
Wife/Husband	0.180***
	(0.017)
Son/Daughter	0.025
	(0.016)
Child-in-law	0.080**
	(0.033)
Grandchild	-0.058
	(0.059)
Father/Mother	0.059
	(0.044)
Brother/Sister	-0.084***
	(0.030)
Parent in law	-0.205
	(0.171)
Nephew/Niece	-0.127*
	(0.073)
Sibling-in-law	-0.037
	(0.074)
Other relative	-0.184**
	(0.082)
Other	-0.455***
	(0.121)
Major Morbidity Days hospitalized (2005)	-0.002**
	(0.001)
Married (2005)	0.042***
	(0.009)
Number of household members (2005)	0.022***
	(0.002)
Age (2005)	0.019***
	(0.002)
Age squared (2005)	-0.000***
	(0.000)
Educ_primary (2005)	-0.036**
	(0.017)
Educ_middle (2005)	0.003
	(0.012)
Educ_secondary (2005)	-0.054***
	(0.014)
Educ_higher (2005)	-0.103***
	(0.013)

Hindu OBC	0.056***
	(0.011)
SCST	0.067***
	(0.012)
Muslim Upper Caste	0.017
	(0.014)
Muslim OBC	0.052**
	(0.022)
Observations	15,452
Pseudo-R^2	7.47
LR chi^2	1512.13
Prob>chi^2	0.000

Source: Author's calculations from the IHDS dataset
Note: Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix 2.7. Occupational mobility estimations without the control function

VARIABLES	Casual-Regular mobility (ref. group: No mobility) Downward	Upward	Industrial mobility (ref. group: No mobility) Mobile	Skill levels in occupations mobility (ref. group: No mobility) Downward	Upward
female	-0.220	-0.025	-0.860***	-0.077	-0.560***
	(0.170)	(0.098)	(0.081)	(0.120)	(0.100)
Hindu OBC	-0.019	-0.135	0.188**	-0.130	-0.106
	(0.183)	(0.101)	(0.081)	(0.123)	(0.092)
SCST	-0.100	-0.541***	0.322***	0.133	-0.016
	(0.223)	(0.124)	(0.095)	(0.137)	(0.111)
Muslim Upper Caste	-0.146	-0.564***	-0.074	-0.297	-0.264
	(0.303)	(0.170)	(0.138)	(0.214)	(0.160)
Muslim OBC	0.106	-0.574***	0.258**	0.202	0.002
	(0.286)	(0.157)	(0.120)	(0.184)	(0.149)
Other	-0.036	-0.288	-0.014	-0.426	-0.245
	(0.334)	(0.211)	(0.156)	(0.261)	(0.176)
Educ_primary (2005)	0.025	0.398***	0.145	0.268	-0.003
	(0.246)	(0.137)	(0.105)	(0.166)	(0.134)
Educ_middle (2005)	0.297*	0.760***	-0.077	0.149	0.135
	(0.160)	(0.098)	(0.075)	(0.125)	(0.092)
Educ_secondary (2005)	0.245	1.231***	-0.047	0.385*	0.183
	(0.367)	(0.182)	(0.148)	(0.234)	(0.177)
Educ_higher (2005)	0.552	1.270***	-0.258	0.731***	0.298
	(0.407)	(0.207)	(0.170)	(0.261)	(0.202)
Sec1 (2005)	-0.025	-0.190	-0.135	0.314	0.309
	(0.407)	(0.199)	(0.165)	(0.248)	(0.188)
Sec2 (2005)	0.044	-0.170	0.094	0.156	0.265
	(0.351)	(0.175)	(0.147)	(0.225)	(0.172)
Sec3 (2005)	0.178	-0.065	-0.120	0.281	-0.039
	(0.403)	(0.202)	(0.170)	(0.257)	(0.202)
Age (2005)	0.237***	-0.014	-0.024	0.085	0.045
	(0.090)	(0.050)	(0.038)	(0.055)	(0.045)
Age squared (2005)	-0.003**	-0.000	0.000	-0.001	-0.000
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Number of children (2005)	-0.075	0.000	-0.041*	-0.085**	-0.047*
	(0.050)	(0.028)	(0.021)	(0.034)	(0.027)
Married (2005)	0.043	-0.182**	-0.187**	-0.141	-0.135
	(0.171)	(0.092)	(0.078)	(0.118)	(0.090)
Selection_correction	1.270*	-0.277	-0.147	0.876**	0.548
	(0.682)	(0.366)	(0.286)	(0.416)	(0.340)
Selection_correction	-0.314	0.324	-0.295	0.091	-0.145
	(0.512)	(0.277)	(0.217)	(0.331)	(0.240)
InitialY	-0.013***	-0.016***	-0.002**	-0.006***	0.004***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
$\widehat{\mu_i}$	-0.006	**-0.061*** **	**-0.004*** **	-0.000	**-0.005*** **

	(0.005)	**(0.004)**	**(0.002)**	(0.003)	**(0.002)**
State control variables	Yes		Yes	Yes	
Constant	-8.641***	0.350	1.198	-3.893**	-2.729**
	(2.575)	(1.386)	(1.076)	(1.550)	(1.270)
Observations	6,789	6,789	6,947	6,947	6,947

Source: Author's calculations from the IHDS dataset
Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix to Chapter 3

Appendix 3.1. Definitions of informality in the Indian legal and institutional framework

System of National Accounts	Informal production units are characterized by: a. Low level of organization b. Little or no division between labor and capital c. Labor relations based on casual employment and/or social relationships as opposed to formal contracts These units belong to the household sector and cannot be associated with other units.
Directorate General of Employment and Training	Employment within the unorganized sector derived as a residual of the total workforce minus the workers in the organized sector.
National Sample Survey Organizaton	a. In the case of manufacturing industries, the enterprises not covered under Annual Survey of Industries constitute the unorganized sector b. In the case of service industries, all enterprises, except for those run by the government (central, state and local body) and in the public sector are regarded as unorganized.
National Commission for Enterprises in the Unorganized Sector	Unorganized sector: « *unorganized sector consists of all unincorporated private enterprises owned by individuals or households engaged in the sale and production of goods and services operated on a proprietary or a partnership basis and with less than ten total workers* ». (and less than 20 without electricity) Unorganized worker: « *Unorganized workers consist of those working in the unorganized sector or households, excluding regular workers with social security benefits provided by the employers and the workers in the formal sector without any employment and social security benefits provided by the employers* »

Source: Lee et al. (2008)

Appendix 3.2. Coefficients of correlation between exclusion restriction variables and earnings variables

	Household business hourly earnings	Number of female infants (<5 y.o.)	Number of male infants (<5 y.o.)	Number of elderly female household members	Number of elderly male household members
Household business hourly earnings	1.000				
Number of female infants (<5 y.o.)	0.039	1.000			
Number of male infants (<5 y.o.)	0.043	0.182	1.000		
Number of elderly female household members	0.094	0.036	0.081	1.000	
Number of elderly male household members	0.009	-0.009	-0.020	0.149	1.000

	Household business hourly earnings	Number of female infants (<5 y.o.)	Number of male infants (<5 y.o.)	Number of elderly female household members	Number of elderly male household members
Individual hourly earnings	1.000				
Number of female infants (<5 y.o.)	-0.062	1.000			
Number of male infants (<5 y.o.)	-0.051	0.181	1.000		
Number of elderly female household members	0.010	0.075	0.080	1.000	
Number of elderly male household members	-0.000	0.041	0.030	0.206	1.000

Source: Author’s calculations from the IHDS dataset

Appendix 3.3. Pairwise comparison of equality of means for social network characteristics

Compared Groups	Social network 1	Social network 2
Hindu OBC – Hindu Upper Castes	-1.027*** (0.049)	-0.836*** (0.059)
SCST – Hindu Upper Castes	-0.832*** (0.054)	-0.924*** (0.065)
Muslim Upper Caste – Hindu Upper Castes	-0.599*** (0.599)	-0.825*** (0.088)
Muslim OBC – Hindu Upper Castes	-1.111*** (0.072)	-1.171*** (0.087)
SCST- Hindu OBC	0.194*** (0.194)	-0.088 (0.063)
Muslim Upper Caste – Hindu OBC	0.427*** (0.082)	0.010 0.098
Muslim Upper Caste-SCST	0.233*** (0.085)	0.087 (0.102)
Muslim OBC- Hindu OBC	-0.084 (0.071)	-0.335*** (0.090)

Source: Author's calculations from the IHDS dataset
Note: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Appendix 3.4. Odds ratios for segment-membership in the salaried sector

	Segment 2	**Segment 3**
Female	16.876***	23.703***
	(0.398)	(0.411)
OBC	0.828***	-0.113
	(0.151)	(0.256)
SCST	0.746***	-0.433
	(0.165)	(0.297)
Upper Caste Muslim	0.679**	-0.372
	(0.268)	(0.436)
OBC Muslim	1.483***	0.180
	(0.240)	(0.412)
Other	0.012	-0.330
	(0.345)	(0.508)
Constant	-0.933***	-0.330***
	(0.145)	(0.508)

Source: Author’s calculations from the IHDS dataset

Appendix to Chapter 4

Appendix 4.1. Quantile decomposition methods

There are two types of quantile decomposition methods emphasized in the literature: the Machado-Mata-Melly decompositions and the RIF-ref decomposition method.

Quantile regressions estimate the marginal effect of the characteristics (independent variables) at each quantile. The method consists in estimating a distribution of log wages for group A individuals being paid on the basis of group B wage structure, and a distribution where group B individuals are paid on the basis of group A wage structure.

Goraus et al. (2017) advise on using the RIF-Reg decomposition approach compared to the MMM one as it provides results that are immune to the path-dependency problem[94] and also "*allows the impact of a particular covariate to separate on the explained and unexplained part of the gap for any quantile of the unconditional distribution of dependent variable*" (Goraus et al. 2017).

The RIF-reg decomposition methodology is a two-step method introduced by Fortin, Lemieux and Firpo (2009, 2011). Using Influence Functions which they define as a function represents the influence of an individual observation on a distributional statistic (for instance the mean or a quantile), the authors introduce the concept of the Recentered Influence Function (RIF) which is equal to the IF to which they add the statistic of interest[95].

The RIF-reg decomposition method simply consists in replacing the dependent variables of the pooled decomposition by the RIF for each quantile of interest (in our case all the deciles d_k). The two following steps are therefore followed for each decile:

1. First, estimating the RIF for each group of interest (e.g. for females and males) and for d_k.
2. Perform the decomposition.

Source: Adapted from Deshpande, Goel, and Khanna (2017) and Firpo (2011)

[94] The path-dependency problem, which causes the decomposition results to be influenced by the order in which the variables are included in the specification, is a common issue in decomposition methods.

[95] We only briefly present the RIF as it calls upon complex statistical concepts that are beyond the scope of this chapter. See Firpo, Fortin & Lemieux (2009) for a detailed explanation of the RIF.

Appendix 4.2. Descriptive Statistics by gender

	Whole sample	Female sample	Male sample
		Mean or percent	
Female	19.98%	N.a.	N.a.
Religion/ Caste			
Hindu Upper Castes	24.56%	22.95%	24.81%
OBC	31.23%	32.12%	31.13%
SCST	27.33%	31.47%	26.37%
Muslim Upper Caste	5.62%	3.70%	6.12%
Muslim OBC	7.78%	5.17%	8.44%
Other	3.46%	4.59%	3.11%
Age	37.821	38.442	37.541
Education level			
None		30.38%	11.00%
Primary	7.08%	5.43%	7.59%
Middle	26.03%	15.34%	28.86%
Secondary	14.78%	8.24%	16.21%
Tertiary	30.51%	32.76%	29.61%
SSC First Class	13.64%	17.04%	12.61%
English	43.21%	39.21%	44.23%
Sector of occupation			
Agriculture	5.25%		
Manufacturing	22.14%	16.23%	23.88%
Services	52.83%	60.83%	50.40%
Public Administration	6.88%	3.22%	7.40%
Construction	12.90%	9.33%	14.01%
Married	70.72%	58.58%	73.60%

Source: Author's calculations from IHDS (2011-12)

Appendix 4.3. Descriptive Statistics by religion and caste group

	Whole sample	Hindu Upper Castes	Hindu OBC	SCST	Muslim upper caste	Muslim OBC
	Mean or percent					
Female	19.98%	18.89%	20.62%	23.10%	13.20%	13.36%
Age	37.821	39.55	37.967	37.430	34.141	34.249
(S.D.)	(11.90)	(11.47)	(11.812)	(11.873)	(12.098)	(12.448)
Education level						
Primary	7.08%	3.15%	7.06%	9.15%	10.51%	10.88%
Middle	26.03%	18.72%	28.23%	28.39%	31.54%	27.26%
Secondary	14.78%	26.77%	16.31%	12.50%	10.39%	10.88%
Tertiary	30.51%	52.83%	27.22%	20.24%	16.75%	13.45%
SSC First Class	13.64%	25.06%	12.88%	6.59%	5.62%	6.46%
English	43.21%	63.88%	41.25%	32.66%	29.22%	27.61%
Sector of occupation						
Agriculture	5.25%					
Manufacturing	22.14%	19.28%	24.60%	18.23%	30.07%	30.97%
Services	52.83%	63.38%	50.92%	46.70%	47.19%	48.67%
Public Administration	6.88%	9.52%	5.37%	8.05%	4.77%	1.50%
Construction	12.90%	5.47%	12.52%	18.91%	15.64%	13.53%
Married	70.72%	72.99%	71.77%	70.73%	61.74%	63.00%

Source: Author's calculations from IHDS (2011-12)

Appendix 4.4. Male-female wage gap across the distribution

Percentile	10	20	30	40	50	60	70	80	90
					MMM decompositions				
	0.508***	0.475***	0.470***	0.461***	0.436***	0.364***	0.320***	0.204***	0.049*
Raw differences	(0.002)	(0.014)	(0.012)	(0.019)	(0.030)	(0.038)	(0.049)	(0.050)	(0.033)
	0.001	0.009	0.012	0.015**	0.017**	0.017***	0.008	0.008	-0.007
Characteristics	(0.011)	(0.009)	(0.009)	(0.009)	(0.011)	(0.013)	(0.015)	(0.019)	(0.016)
	0.510***	0.465	0.458***	0.446***	0.418***	0.377***	0.195***	0.195***	-0.056***
Coefficients	(0.002)	(0.012)	(0.008)	(0.006)	(0.005)	(0.003)	(0.003)	(0.006)	(0.011)
					RIF-Reg decompositions				
Estimated Male wage from RIF-regression	2.379***	2.625***	2.776***	3.002***	3.133***	3.329***	3.546***	3.856***	4.216***
	(0.009)	(0.011)	(0.010)	(0.010)	(0.012)	(0.012)	(0.015)	(0.013)	(0.015)
Estimated Female wage from RIF-regression	1.584***	1.880***	2.044***	2.172***	2.335***	2.470***	2.711***	3.108***	3.581***
	(0.023)	(0.021)	(0.020)	(0.021)	(0.024)	(0.027)	(0.033)	(0.040)	(0.044)
Difference	0.795***	0.744***	0.732***	0.830***	0.798***	0.859***	0.835***	0.748***	0.635***
	(0.025)	(0.023)	(0.024)	(0.024)	(0.027)	(0.030)	(0.037)	(0.042)	(0.047)
Explained	-0.004	0.004	-0.002	0.014	0.023*	0.028*	0.045***	0.055***	0.026
	(0.013)	(0.012)	(0.011)	(0.012)	(0.014)	-0.015	(0.018)	(0.017)	(0.022)
Unexplained	0.800***	0.741***	0.734***	0.816***	0.776***	0.831***	0.789***	0.693***	0.609***
	(0.026)	(0.024)	(0.024)	(0.023)	(0.027)	-0.029	(0.035)	(0.038)	(0.048)
Observations	50,129	50,129	50,129	50,129	50,129	50,129	50,129	50,129	50,129

Source: Author's calculations from IHDS (2011-12)
Note: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Appendix 4.5. Non-parametric decomposition results by religion and caste

	OBCs compared to the other groups					
	Coefficient	Percent of the gap	% matched in ref. group	% of matched in rest of sample	Wage gap	Wage gap in the matched sample
Model 1	***age and highest educational attainment***					
Δ	-0.023	(-) 100 %	100	99.67	-0.079***	-0.059***
Δ_0 (std. error)	-0.018 (0.003)	(-) 78.2 %				
Δ_X	-0.006	(-) 26.1 %				
Δ_A	0.000	0 %				
Δ_B	0.001	4.3 %				
Model 2	*Age, highest educational attainment and **SSC class (none, 1, 2 or 3)***					
Δ	-0.023	(-) 100 %	99.50	98.49	0.079***	-0.077***
Δ_0 (std. error)	-0.019 (0.003)	(-) 82.6 %				
Δ_X	-0.004	(-) 17.4 %				
Δ_A	0.000	0				
Δ_B	0.000	0 %				
Model 3	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), **sector of occupation (5 category)***					
Δ	-0.023	(-) 100 %	94.55	88.75	0.079***	-0.074***
Δ_0 (std. error)	-0.016 (0.001)	(-) 69.6 %				
Δ_X	-0.006	(-) 26 %				
Δ_A	0.001	4.3 %				
Δ_B	-0.002	(-) 8.7 %				
Model 4	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 category), **casual occupation***					
Δ	-0.023	(-) 100 %	90.15	81.92	0.079***	0.075***
Δ_0	-0.013 (0.000)	(-) 56.5 %				
Δ_X	-0.009	(-) 39.1 %				
Δ_A	-0.000	(-) 0 %				
Δ_B	-0.001	4.3 %				
Model 5	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 category), casual occupation, **type of occupation (8 skill categories)***					
Δ	-0.023	(-) 100 %	71.36	56.51	0.079***	0.059**
Δ_0	-0.018 (0.000)	(-) 78.3 %				
Δ_X	-0.002	(-) 8.7 %				
Δ_A	0.005	21.7 %				
Δ_B	-0.008	(-) 34.8 %				
Model 6	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 category), casual occupation, type of occupation and **gender***					
Δ	-0.023	(-) 100 %	64.80	49.25	0.079***	0.036**
Δ_0	-0.011 (0.000)	(-) 47.8 %				
Δ_X	-0.001	(-) 4.3 %				
Δ_A	-0.003	(-) 13 %				
Δ_B	-0.008	(-) 34.8 %				

Source: Author's calculations from IHDS (2011-12)

	SCST compared to other groups					
	Coefficient	Percent of the gap	% matched in ref. group	% of matched in rest of sample	Wage gap	Wage gap in the matched sample
Model 1	***age and highest educational attainment***					
Δ	-0.032	(-) 100 %	100	99.31	-0.108***	-0.111***
Δ_0 (std. error)	0.015 (0.002)	46.9 %				
Δ_X	-0.048	(-) 150 %				
Δ_A	0.000	0 %				
Δ_B	0.001	3.1 %				
Model 2	*Age, highest educational attainment and* ***SSC class (none, 1, 2 or 3)***					
Δ	-0.032	(-) 100 %	99.68	96.59	-0.108***	-0.103***
Δ_0 (std. error)	0.023 (0.003)	71.9 %				
Δ_X	-0.053	(-) 165.6 %				
Δ_A	0.000	0 %				
Δ_B	-0.002	(-) 6.2 %				
Model 3	*Age, highest educational attainment, SSC class (none, 1, 2 or 3),* ***sector of occupation (5 category)***					
Δ	-0.032	(-) 100 %	93.81	85.22	-0.108***	-0.092***
Δ_0 (std. error)	0.014 (0.001)	43.7 %				
Δ_X	-0.041	(-) 128.1 %				
Δ_A	0.002	6.2 %				
Δ_B	-0.007	(-) 21.9 %				
Model 4	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 category),* ***casual occupation***					
Δ	-0.032	(-) 100 %	90.18	77.77	-0.108***	-0.098***
Δ_0	0.013 (0.000)	40.6 %				
Δ_X	-0.042	(-) 131.2 %				
Δ_A	0.002	6.2 %				
Δ_B	-0.005	(-) 15.6 %				
Model 5	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 category), casual occupation,* ***type of occupation (8 skill categories)***					
Δ	-0.032	(-) 100 %	72.56	51.22	-0.108***	-0.067***
Δ_0	0.012 (0.000)	37.5 %				
Δ_X	-0.031	(-) 96.9 %				
Δ_A	0.004	12.5 %				
Δ_B	-0.017	(-) 53.1 %				
Model 6	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 category), casual occupation, type of occupation and* ***gender***					
Δ	-0.032	(-) 100 %	64.42	43.95	-0.108***	-0.061***
Δ_0	0.007 (0.000)	21.9 %				
Δ_X	-0.027	(-) 84.4 %				
Δ_A	0.002	6.2 %				
Δ_B	-0.014	(-) 43.7 %				

Source: Author's calculations from IHDS (2011-12)

	Upper Caste Muslim compared to other groups					
	Coefficient	Percent of the gap	% matched in ref. group	% of matched in rest of sample	Wage gap	Wage gap in the matched sample
Model 1	***age and highest educational attainment***					
Δ	-0.077	(-) 100 %	100	92.59	-0.259***	-0.244***
Δ_0 (std. error)	-0.018 (0.002)	(-) 23.4 %				
Δ_X	-0.054	(-) 70.1 %				
Δ_A	-0.000	(-) 0 %				
Δ_B	-0.005	(-) 6.5 %				
Model 2	*Age, highest educational attainment and **SSC class (none, 1, 2 or 3)***					
Δ	-0.077	(-) 100 %	99.63	78.96	-0.259***	-0.166***
Δ_0 (std. error)	-0.013 (0.002)	(-) 16.9 %				
Δ_X	-0.037	(-) 48 %				
Δ_A	0.001	1.3 %				
Δ_B	-0.028	(-) 36.4 %				
Model 3	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), **sector of occupation (5 category)***					
Δ	-0.077	(-) 100 %	97.11	52.20	-0.259***	-0.157***
Δ_0 (std. error)	-0.008 (0.002)	(-) 10.4 %				
Δ_X	-0.038	(-) 49.3 %				
Δ_A	0.001	1.3 %				
Δ_B	-0.032	(-) 41.5 %				
Model 4	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 category), **casual occupation***					
Δ	-0.077	(-) 100 %	94.81	41.97	-0.259***	-0.194***
Δ_0	-0.007 (0.000)	(-) 9.1 %				
Δ_X	-0.051	(-) 66.2 %				
Δ_A	-0.000	(-) 0 %				
Δ_B	-0.019	(-) 24.7 %				
Model 5	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 category), casual occupation, **type of occupation (8 skill categories)***					
Δ	-0.077	(-) 100 %	79.64	24.41	-0.259***	-0.136***
Δ_0	-0.023 (0.000)	(-) 29.9 %				
Δ_X	-0.020	(-) 26 %				
Δ_A	-0.000	(-) 0 %				
Δ_B	-0.034	(-) 44.1 %				
Model 6	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 category), casual occupation, type of occupation and **gender***					
Δ	-0.077	(-) 100 %	76.14	19.47	-0.259***	-0.140***
Δ_0	-0.029 (0.000)	(-) 37.7 %				
Δ_X	-0.012	(-) 15.6 %				
Δ_A	-0.003	(-) 8.1 %				
Δ_B	-0.032	(-) 41.5%				

Source: Author's calculations from IHDS (2011-12)

	OBC Muslim compared other groups					
	Coefficient	Percent of the gap	% matched in ref. group	% of matched in rest of sample	Wage gap	Wage gap in the matched sample
Model 1	***age and highest educational attainment***					
Δ	-0.096	(-) 100 %	100	91.90	-0.326***	-0.286***
Δ_0 (std. error)	-0.014 (0.002)	(-) 14.6 %				
Δ_X	-0.070	(-) 72.9 %				
Δ_A	0.000	0 %				
Δ_B	-0.012	(-) 12.5 %				
Model 2	*Age, highest educational attainment and **SSC class (none, 1, 2 or 3)***					
Δ	-0.096	(-) 100 %	99.74	82.64	-0.326***	-0.244***
Δ_0 (std. error)	-0.012 (0.002)	(-) 12.5 %				
Δ_X	-0.061	(-) 63.5 %				
Δ_A	0.000	0 %				
Δ_B	-0.024	(-) 25 %				
Model 3	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), **sector of occupation (5 category)***					
Δ	-0.096	(-) 100 %	97.13	59.72	-0.326***	-0.197***
Δ_0 (std. error)	-0.013 (0.001)	(-) 13.5 %				
Δ_X	-0.045	(-) 46.9 %				
Δ_A	-0.001	(-) 1 %				
Δ_B	-0.037	(-) 38.5 %				
Model 4	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 category), **casual occupation***					
Δ	-0.096	(-) 100 %	94.25	48.62	-0.326***	-0.200***-
Δ_0	-0.015 (0.000)	(-) 15.6 %				
Δ_X	-0.044	(-) 45.8 %				
Δ_A	-0.001	(-) 1%				
Δ_B	-0.036	(-) 37.5 %				
Model 5	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 category), casual occupation, **type of occupation (8 skill categories)***					
Δ	-0.096	(-) 100 %	80.05	28.48	-0.326***	0.143***
Δ_0	-0.015 (0.000)	(-) 15.6 %				
Δ_X	-0.029	(-) 30.2 %				
Δ_A	-0.004	(-) 4.2 %				
Δ_B	-0.048	(-) 50 %				
Model 6	*Age, highest educational attainment, SSC class (none, 1, 2 or 3), sector of occupation (5 category), casual occupation, type of occupation and **gender***					
Δ	-0.096	(-) 100 %	75.70	22.70	-0.326***	-0.144***
Δ_0	-0.023 (0.000)	(-) 23.9 %				
Δ_X	-0.022	(-) 22.9%				
Δ_A	-0.008	(-) 8.3 %				
Δ_B	-0.044	(-) 45.8 %				

Source: Author’s calculations from IHDS (2011-12)

Appendix 4.6. Quantile decomposition results (Hindu Upper Caste compared to other groups)

Percentile	10	20	30	40	50	60	70	80	90
	MMM decompositions								
	-0.508***	-0.475***	-0.470***	-0.461***	-0.436***	-0.364***	-0.320***	-0.204***	-0.049*
Raw differences	(0.002)	(0.014)	(0.012)	(0.019)	(0.030)	(0.038)	(0.049)	(0.050)	(0.033)
	0.001	-0.009	-0.012	-0.015**	-0.017**	-0.017***	-0.008	-0.008	0.007
Characteristics	(0.011)	(0.009)	(0.009)	(0.009)	(0.011)	(0.013)	(0.015)	(0.019)	(0.016)
	-0.510***	-0.465	-0.458***	-0.446***	-0.418***	-0.377***	-0.195***	-0.195***	0.056***
Coefficients	(0.002)	(0.012)	(0.008)	(0.006)	(0.005)	(0.003)	(0.003)	(0.006)	(0.011)
	RIF-Reg decompositions								
Estimated Male wage from RIF-regression	2.379***	2.625***	2.776***	3.002***	3.133***	3.329***	3.546***	3.856***	4.216***
	(0.009)	(0.011)	(0.010)	(0.010)	(0.012)	(0.012)	(0.015)	(0.013)	(0.015)
Estimated Female wage from RIF-regression	1.584***	1.880***	2.044***	2.172***	2.335***	2.470***	2.711***	3.108***	3.581***
	(0.023)	(0.021)	(0.020)	(0.021)	(0.024)	(0.027)	(0.033)	(0.040)	(0.044)
Difference	0.795***	0.744***	0.732***	0.830***	0.798***	0.859***	0.835***	0.748***	0.635***
	(0.025)	(0.023)	(0.024)	(0.024)	(0.027)	(0.030)	(0.037)	(0.042)	(0.047)
Explained	-0.004	0.004	-0.002	0.014	0.023*	0.028*	0.045***	0.055***	0.026
	(0.013)	(0.012)	(0.011)	(0.012)	(0.014)	-0.015	(0.018)	(0.017)	(0.022)
Unexplained	0.800***	0.741***	0.734***	0.816***	0.776***	0.831***	0.789***	0.693***	0.609***
	(0.026)	(0.024)	(0.024)	(0.023)	(0.027)	-0.029	(0.035)	(0.038)	(0.048)
Observations	50,129	50,129	50,129	50,129	50,129	50,129	50,129	50,129	50,129

Source: Author's calculations from IHDS (2011-12)

Résumé en français

Le développement économique et l'égalité entre les groupes composant la société sont complémentaires (The World Bank 2012). Dans ce cadre, il est nécessaire que les politiques publiques soient en mesure d'assurer l'égalité des opportunités de chacun. La persistance des inégalités socioéconomiques entre groupes de genre, de caste et de religion reflète une stratification importante de la société indienne. En effet, l'Inde est classée 131ème pays en termes de développement humain par sexe. Les discriminations envers les femmes sont présentes dans tous les domaines de la société avec, par exemple, une préférence marquée pour les fils et des phénomènes d'avortement sélectif. Parmi les groupes socio-religieux, des groupes issus du système de castes sont organisés de manière hiérarchique. Les *Hindu Upper Castes* (Hautes Castes Hindoues) sont les plus avantagés. Les *Scheduled Castes and Scheduled Tribes ou SCST* (les plus basses castes) sont répertoriées dans la Constitution indienne et bénéficient de politiques de discrimination positive. Un groupe intermédiaire, *Other Backward Castes ou OBC* (le « autres basses castes ») bénéficie également de politiques de discrimination positive mais moins importante que les *SCST*. Par ailleurs, les Musulmans faisant partie de la société indienne sont aussi divisés en deux groupes reflétant une structure de caste, les *Muslim Upper Castes* représente le groupe qui est économiquement le plus avantagé et les *Muslim OBC* représente le groupe désavantagé, faisant lui aussi l'objet de politiques de discrimination positives. L'objet de cette thèse est d'explorer l'existence et l'ampleur des inégalités horizontales fondées sur le genre, la religion et la caste sur le marché du travail urbain en Inde.

Quatre questionnements spécifiques sont abordés pour permettre de détecter l'existence d'inégalités horizontales et pour en comprendre l'étendue et les mécanismes sous-jacents. Nous utilisons pour cela des méthodologies quantitatives et la base de données *India Human Development Survey* (Desai, Vanneman, and National Council of Applied Economic Research 2012). Des études de cas, issues d'une étude de terrain dans la ville de Ranipet (Tamil Nadu) illustrent ou nuancent les résultats des études quantitatives.

Un premier chapitre met en évidence les liens entre les disparités des groupes (en termes de santé et d'éducation) et l'exclusion du marché du travail. En définissant l'exclusion du marché du travail

comme l'union de l'inactivité et du chômage, les résultats d'une estimation logistique multinomiale montrent que les femmes sont particulièrement propices à l'exclusion. En effet, la probabilité pour une femme d'être exclue du marché du travail est particulièrement élevée, et ce pour tous les niveaux d'éducation. Pour les hommes, la probabilité de participer au marché du travail est bien plus importante que la probabilité d'être exclu. Dans un deuxième temps, nous explorons la relation intergénérationnelle entre le statut de la mère sur le marché du travail et l'éducation des enfants. Une originalité de cette étude est d'opérer une distinction entre les femmes qui travaillent, les femmes qui ne travaillent pas mais qui ont le droit de travailler et les femmes qui n'ont pas le droit de travailler. L'existence de l'exclusion forcée du marché du travail est en effet une réalité pour un nombre considérable de femmes en Inde (Miller 1982). En utilisant une méthode de scores de propension adaptée aux variables multinomiales et en considérant la participation sur le marché du travail de la mère comme une variable de traitement, nous mettons en évidence un résultat paradoxal. La participation de la mère sur le marché du travail contribue à élargir les écarts de genre en matière d'éducation des enfants. En effet, le fait que les femmes aient le droit de travailler est associé à une baisse du score des filles alors qu'il est associé à une hausse ou à aucun effet sur le score des garçons. Les filles qui sont issues des ménages où le travail des femmes est stigmatisé ont en effet des meilleures notes, mettant ainsi en exergue le paradoxe suivant. Elles sont les plus équipées pour le marché du travail mais seront probablement, à leur tour, interdites de travailler lorsqu'elles auront atteint l'âge adulte. Avoir la permission de travailler n'affecte pas significativement le temps passé à l'école et le nombre de jour d'absence. De plus, avoir un emploi à temps-plein affecte négativement les heures passées à faire les devoirs pour les garçons et pour les filles. Nous supposons alors que pour que le travail des femmes contribue à diminuer les écarts de genre en matière d'éducation, il est nécessaire que ce travail puisse être qualifié de « *empowering* » et qu'il ne soit pas exclusivement un moyen de subsistance. En effet, cela ne permet pas de créer un effet de motivation chez les filles.

Le deuxième chapitre s'intéresse à la mobilité occupationnelle et de revenus entre 2005 and 2011-12. Nous utilisons la dimension de panel de la base de données pour observer les transitions entre occupations et la mobilité relative au sein de la distribution des revenus. Afin de détecter la mobilité professionnelle, nous mesurons les transitions entre les emplois occasionnels et réguliers, les industries et le niveau de compétence requis des professions. Nous détections également les changements de percentiles au sein de la distribution du revenu horaire entre les deux dates. Les résultats montrent que les femmes sont plus immobiles que les hommes en termes de profession, leur mobilité relative en termes de revenus n'est pas significativement différente de celle des hommes. Les castes supérieures hindoues (*Hindu Upper Castes*) présentent un taux plus élevé de personnes qui migrent vers des emplois hautement qualifiés et vers le secteur des services. Leur mobilité en termes de revenus horaires est toutefois inférieure à celle des autres groupes socio-religieux. Ceci suggère un processus de « *rattrapage* ». Nous estimons ensuite les déterminants de la mobilité en considérant plusieurs sources potentielles de biais (le biais de sélection lié à la participation au marché du travail et celui lié à l'attrition ; les erreurs de mesures dans la déclaration des revenus ; le biais d'endogénéité lié à la variable du revenu de 2005). Les résultats montrent que le niveau d'éducation est un facteur déterminant de la mobilité professionnelle. Néanmoins, des différences significatives subsistent entre les groupes socio-religieux, toutes choses égales par ailleurs. Comparativement aux castes supérieures hindoues, la probabilité de mobilité vers des professions de meilleure qualité (c'est-à-dire un emploi régulier ou avec un degré de compétences plus élevé) est plus faible pour les *SCST*. Les musulmans ont également beaucoup moins de chances de retrouver un emploi régulier. L'appartenance à un groupe de castes spécifique n'affecte pas de manière significative les chances de mobilité ascendante, sauf au sommet de la distribution où les hautes castes musulmanes ont des chances de mobilité significativement plus faibles. Deux tendances opposées affectent les femmes. Leur faible niveau d'éducation limite leur transition vers des emplois plus qualifiés, mais toutes choses égales par ailleurs, elles bénéficient d'une probabilité plus élevée de mobilité professionnelle que les hommes. Les résultats montrent que la tendance de rattrapage des groupes défavorisés passe par l'éducation. Cette étude étant effectuée sur des revenus horaires, il n'est pas exclu que l'accès à un nombre d'heures de travail supplémentaire ou à des transferts puissent être une source de mobilité économique.

Le troisième chapitre analyse l'existence d'une segmentation du marché du travail dans une économie essentiellement informelle. En utilisant une méthode semi-paramétrique de modèle de mélange fini permettant de détecter la présence de segments, nous montrons que le secteur des entreprises familiales (au sein duquel se trouvent les travailleurs indépendants) a une structure homogène, tandis que le secteur du salariat est segmenté. Cette segmentation indique une forte ségrégation genrée sur le marché du travail. En effet, les femmes appartiennent à un segment distinct du marché du travail avec un salaire moyen inférieur à celui des deux autres segments combinés. Les hommes sont divisés en deux segments : un segment supérieur et segment inférieur.

Nous montrons que le segment des femmes et le segment inférieur des hommes sont des trappes au sein du marché du travail. L'informalité étant prédominante dans l'Inde urbaine, les emplois de meilleure qualité sont concentrés dans le segment supérieur des hommes, mais nous ne trouvons aucune indication d'un clivage formel - informel. Les femmes et l'ensemble des groupes de caste et de religion font face à des obstacles pour accéder à ce segment. Pour conclure ce chapitre met en évidence le phénomène de ségrégation sur le marché du travail indien.

Le quatrième chapitre analyse les sources d'écarts de salaire en comparant les résultats de décompositions paramétriques et non-paramétriques. Nous constatons que l'écart salarial entre hommes et femmes n'est pas dû à une *discrimination salariale pure*. L'effet de la sélection par la profession et la ségrégation dans différents types de professions constituent la principale source des écarts de salaire entre les femmes et les hommes. L'écart de salaire entre les hautes castes hindoues et le reste de la population est en partie dû au népotisme, à la discrimination et aux différences de caractéristiques des groupes. Ce sont les *OBC* qui subissent la discrimination alors que les *SCST* et les musulmans ont des écarts de salaires qui sont principalement liés à leurs caractéristiques.

Bibliography

Afridi, Farzana, Abhiroop Mukhopadhyay, and Soham Sahoo. 2016. “Female Labor Force Participation and Child Education in India: Evidence from the National Rural Employment Guarantee Scheme.” *IZA Journal of Labor and Development* 5 (1): 7.

Agrawal, Tushar. 2014. “Gender and Caste-Based Wage Discrimination in India: Some Recent Evidence.” *Journal for Labour Market Research* 47 (4): 329–40.

Akee, Randall. 2011. “Errors in Self-Reported Earnings: The Role of Previous Earnings Volatility and Individual Characteristics.” *Journal of Development Economics* 96 (2): 409–21.

Akerlof, George A., and Rachel E. Kranton. 2000. “Economics and Identity*.” *Quarterly Journal of Economics* 115 (3): 715–53.

Alcaraz, Carlo, Daniel Chiquiar, and Alejandrina Salcedo. 2015. “Informality and Segmentation in the Mexican Labor Market.” *Bank of Mexico Working Paper* 2015–25.

Alkon, Meir. 2018. “Do Special Economic Zones Induce Developmental Spillovers? Evidence from India’s States.” *World Development* 107: 396–409.

Altonji, Joseph G., Anthony A. Smith, and Ivan Vidangos. 2009. “Modeling Earnings Dynamics.” *NBER Working Paper No. w14743*

Altonji, Joseph J, and Rebecca M Blank. 1999. “Race and Gender in the Labor Market.” In *Handbook of Labor Economics*, 3:3143–3259.

Amelot, Xavier, and Loraine Kennedy. 2010. “Dynamique économique et recompositions territoriales, une industrie traditionnelle locale de l’Inde du sud face à la mondialisation.” *Annales de géographie* 671–672 (1): 137.

Angrist, Joshua David, and Jörn-Steffen Pischke. 2009. “Mostly Harmless Econometrics : An Empiricist’s Companion.” In , 373. Princeton University Press.

Arrow, Kenneth J. 1973. “The Theory of Discrimination.” In *Discrimination in Labor Markets*, edited by Orley Ashenfelter and Rees Albert. Princeton University Press.

Arulampalam, Wiji, and Alison L. Booth. 1998. “Training and Labour Market Flexibility: Is There a Trade-off? *.” *British Journal of Industrial Relations* 36 (4): 521–36.

Asfaw, Abay, Francesca Lamanna, and Stephan Klasen. 2010. “Gender Gap in Parents’ Financing Strategy for Hospitalization of Their Children: Evidence from India.” *Health Economics* 19 (3): 265–79.

Asian Development Bank. 2018. “ADB Basic 2018 Statistics,” www.adb.org/publications/basic-statistics-2018

Azam, Mehtabul. 2016. “Household Income Mobility in India, 1993-2011.” *SSRN Electronic Journal*, no. 10308: 1993–2011.

Azam, Mehtabul, Aimee Chin, and Nishith Prakash. 2013. “The Returns to English-Language Skills in India.” *Economic Development and Cultural Change* 61 (2): 335–67.

Azam, Mehtabul, and Geeta Gandhi Kingdon. 2013. “Are Girls the Fairer Sex in India? Revisiting Intra-Household Allocation of Education Expenditure.” *World Development* 42 (1): 143–64.

Babcock, Linda, and Sara Laschever. 2003. *Women Don’t Ask : Negotiation and the Gender Divide*. Princeton University Press.

Banerjee, Abhijit, Marianne Bertrand, Saugato Datta, and Sendhil Mullainathan. 2009. “Labor Market Discrimination in Delhi: Evidence from a Field Experiment.” *Journal of Comparative Economics* 37 (1): 14–27.

Bargain, Olivier, and Prudence Kwenda. 2011. “Earnings Structures, Informal Employment, And Self-Employment: New Evidence From Brazil, Mexico, And South Africa.” *Review of Income and Wealth* 57 (SUPPL. 1): 100–122.

Battisti, Michele. 2013. “Reassessing Segmentation in the Labour Market: An Application for Italy 1995-2004.” *Bulletin of Economic Research* 65 (S1): 38–55.

Beauvoir, Simone de. 1949. *Le Deuxième Sexe*. Ed. 2002 Gallimard.

Becker, Gary S. 1962. “Investment in Human Capital: A Theoretical Analysis.” *Journal of Political Economy* 70 (5, Part 2): 9–49.

Becker, Gary S. 1971. *The Economics of Discrimination*. The University of Chigaco Press.

Behrman, Jere R, and Anil B Deolalikar. 1990. “The Intrahousehold Demand for Nutrients in Rural South India: Individual Estimates, Fixed Effects, and Permanent Income.” *The Journal of Human Resources* 25 (4): 665–96.

Benbabaali, Dalal. 2013. *Caste Dominante et Territoire En Inde Du Sud : Migration et Ascension Sociale Des Kamma d’ Andhra Côtier*. Université Paris Ouest Nanterre La Défense.

Bergmann, Barbara R. 1974. “Occupational Segregation, Wages and Profits When Employers Discriminate by Race or Sex.” *Source: Eastern Economic Journal* 1 (2): 103–10.

Bertrand, Marianne, Chang-Tai Hsieh, and Nick Tsivanidis. 2017. “Contract Labor and Firm Growth in India.”, Mimeo

Bharadwaj, Prashant, Giacomo De Giorgi, David Hansen, and Christopher A. Neilson. 2016. “The Gender Gap in Mathematics: Evidence from Chile.” *Economic Development and Cultural Change* 65 (1): 141–66.

Bhattacharya, Debopam, and Bhashkar Mazumder. 2011. “A Nonparametric Analysis of Black-White Differences in Intergenerational Income Mobility in the United States.” *Quantitative Economics* 2 (3): 335–79.

Bhattacharya, Gargi, and Sushil Kr Haldar. 2015. “Does Demographic Dividend Yield Economic Dividend ? India , a Case Study.” *Economics Bulletin* 35 (2): 1274–91.

Bhaumik, Sumon Kumar, and Manisha Chakrabarty. 2009. “Is Education the Panacea for Economic Deprivation of Muslims?” *Journal of Asian Economics* 20 (2): 137–49.

Blau, Francine D, and Lawrence M Kahn. 2017. “Changes in the Labor Supply Behavior of Married Women : 1980 – 2000” 25 (3): 393–438.

Boeri, Tito, and Jan van Ours. 2013. *The Economics of Imperfect Labor Markets*. *Second Edition*. Princeton University Press.

Boillot, Jean-Joseph. 2016. *L’économie de l’Inde*. La Découverte.

Borjas, George. 2010. “Labor Supply.” In *Labor Economics*, 21–83.

Borjas, George, 2012. “Labor Market Equilibrium.” In *Labor Economics*, 144–202.

Borooah, Vani K. 2012. “Inequality in Health Outcomes in India: The Role of Caste and Religion.”

In *Blocked Blocked by Caste Economic Discrimination in Modern India*, edited by Thorat Sukhadeo and S. Neuman Katherine. Oxford University Press.

BOSERUP, E. 1970. *Women's Role in Economic Development*. Earthscan. 1970.

Bourguignon, François, Martin Fournier, and Marc Gurgand. 2007. "Selection bias corrections based on the multinomial logit model: monte carlo comparisons." *Journal of Economic Surveys* 21 (1): 174–205.

Bros, Catherine. 2010. *Castes in India : Implications of Social Identity In Economics*. Université Paris I Panthéon-Sorbonne

Brunetti, Irene and Fiaschi, Davide, (2015), "Intragenerational Mobility in Italy: a Non-parametric Estimates", *Dipartimento di Economia e Management (DEM) Discussion Papers*, University of Pisa, Pisa, Italy,

Buchinsky, Moshe, and Jennifer Hunt. 1999. "Wage Mobility in the United States." *Review of Economics and Statistics* 81 (3): 351–68.

Burkhauser, Richard V., and Kenneth A. Couch. 2012. "Intragenerational Inequality and Intertemporal Mobility." In *The Oxford Handbook of Economic Inequality*, edited by W. Salverda, B. Nolan, and T.M. Smeeding, 522–48. Oxford University Press.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics : Methods and Applications*. Cambridge University Press.

Carneiro, Pedro, James J. Heckman, and Dimitriy V. Masterov. 2005. "Labor Market Discrimination and Racial Differences in Premarket Factors." *The Journal of Law and Economics* 48 (1): 1–39.

Castells, Manuel, and Alejandro Portes. 1989. "World Underneath: The Origins, Dynamics and Effects of the Informal Economy." In *The Informal Economy: Studies in Advanced and Less Developed Countries*, edited by Alejandro Portes, Manuel Castells, and Lauren A. Benton. John Hopkins University Press.

Cattaneo, Matias D. 2010. "Efficient Semiparametric Estimation of Multi-Valued Treatment Effects under Ignorability." *Journal of Econometrics* 155 (2): 138–54.

Cazes, Sandrine, and Sher Verick. 2013. *Perspectives on Labour Economics for Development*. Edited by Sandrine Cazes and Sher Verick. Geneva: ILO.

Chakraborty, Tanika, Anirban Mukherjee, Swapnika Reddy Rachapalli, and Sarani Saha. 2018. “Stigma of Sexual Violence and Women’s Decision to Work.” *World Development* 103: 226–38.

Chamarbagwala, Rubiana. 2006. “Economic Liberalization and Wage Inequality in India.” *World Development* 34 (12): 1997–2015.

Chaudhary, Ruchika, and Sher Verick. 2014. “Female Labour Force Participation in India and Beyond.” *International Labour Organization*. Geneva: ILO.

Chen, Martha Alter. 2006. “Rethinking the Informal Economy: Linkages with the Formal Economy and the Formal Regulatory Environment.” In *Linking the Formal and Informal Economy*, 75–92. Oxford University Press.

Chen, Marty, and Jean Drèze. 1992. “Widows and Health in Rural North India.” *Economic and Political Weekly* 27 (43): WS81-WS92.

Cigno, Alessandro., and Furio C. Rosati. 2005. *The Economics of Child Labour*. Oxford University Press.

Coate, S, and G C Loury. 1993. “Will Affirmative-Action Policies Eliminate Negative Stereotypes.” *American Economic Review* 83 (5): 1220–40.

Corak, Miles. 2013. “Income Inequality, Equality of Opportunity, and Intergenerational Mobility.” *Journal of Economic Perspectives* 27 (3): 79–102.

Cotton, Jeremiah. 1988. “On the Decomposition of Wage Differentials.” *The Review of Economics and Statistics* 70 (2): 236–43.

Covarrubias, Arlette. 2013. “Social Norms and Women’s Participation in Salaried Employment : The Case of the Tehuacan Region of Mexico” *Bulletin of Latin American Research* 32 (1): 17–31.

Crenshaw, K. (1997). Beyond racism and misogyny: Black feminism and 2 Live Crew. In C. J. Cohen, K. B. Jones, & J. C. Tronto (Eds.), Women transforming politics: An alternative reader (pp. 549–68). New York: New York University Press.

Crespo Cuaresma, Jesús, Wolfgang Lutz, and Warren Sanderson. 2014. "Is the Demographic Dividend an Education Dividend?" *Demography* 51 (1): 299–315.

Crespo, Nuno, Nadia Simoes, and Sandrina B. Moreira. 2014. "Gender Differences in Occupational Mobility - Evidence from Portugal." *International Review of Applied Economics* 28 (4): 460–81.

Dang, Hai-Anh H., and Peter F. Lanjouw. 2018. "Poverty Dynamics in India between 2004 and 2012: Insights from Longitudinal Analysis Using Synthetic Panel Data." *Economic Development and Cultural Change 67(1):131-170*.

Deininger, Klaus, Songqing Jin, and Hari Nagarajan. 2013. "Wage Discrimination in India's Informal Labor Markets: Exploring the Impact of Caste and Gender." *Review of Development Economics* 17 (1): 130–47.

Deliège, Robert. 2004. *Les Castes En Inde Aujourd'hui*. Presses universitaires de France.

Deliège, Robert. 2004. "Chapitre 2. La caste selon Louis Dumont." In *Les castes en Inde aujourd'hui*, 45–72. Sociologie d'aujourd'hui. Paris cedex 14: Presses Universitaires de France.

Denis, Eric, and Kamala Marius-Gnanou. 2010. "Toward a Better Appraisal of Urbanization in India." *Cybergeo : European Journal of Geography*, November 2010.

Desai, Sonalde, Reeve Vanneman, and National Council of Applied Economic Research. 2012. "India Human Development Survey-II (IHDS-II)." https://ihds.umd.edu/

Deshpande, Ashwini. 2000. "Does Caste Still Define Disparity? A Look at Inequality in Kerala, India." *American Economic Review* 90 (2): 322–25.

Deshpande, Ashwini, Deepti Goel, and Shantanu Khanna. 2017. "Bad Karma or Discrimination? Male–Female Wage Gaps Among Salaried Workers in India." *World Development* 102: 331-344 (August).

Deshpande, Ashwini, and Smriti Sharma. 2016. "Disadvantage and Discrimination in Self-Employment: Caste Gaps in Earnings in Indian Small Businesses." *Small Business Economics* 46 (2): 325–46.

Dhar, Diva, Tarun Jain, and Seema Jayachandran. 2015. “Intergenerational Transmission of Gender Attitudes: Evidence from India.” *NBER Working Paper Series*.

Dickens, William T, and Kevin Lang. 1985. “A Test of Dual Labor Market Theory.” *American Economic Review* 75 (4): 792–805.

Dimou, Michel. 2006. “J.-E. Cairnes : Groupes Non Concurrents et Organisation Industrielle.” *Revue d’économie Industrielle*, no. 113 (March): 31–44.

Dimova, Ralitza, Christophe J. Nordman, and François Roubaud. 2010. “Allocation of Labor in Urban West Africa: Insights from the Pattern of Labor Supply and Skill Premiums.” *Review of Development Economics* 14 (1): 74–92.

Dirks, Nicholas B. 2001. “Introduction: The Modernity of Caste.” *Caste of Mind:Colonialism and the Making of Modern India*, 3–18.

Doeringer, Peter B., and Michael J. Piore. 1985. *Internal Labor Markets and Manpower Analysis*. M.E. Sharpe.

Drèze, Jean, and Amartya Sen. 1999. *India: Economic Development and Social Opportunity*. Oxford University Press.

Drèze, Jean, and Amartya Sen. 2013. *An Uncertain Glory : India and Its Contradictions*. Princeton University Press.

Dumas, Christelle. 2012. “Does Work Impede Child Learning? The Case of Senegal.” *Economic Development and Cultural Change* 60 (4): 773–93.

Eagly, Alice H., and Wendy Wood. 1999. “The Origins of Sex Differences in Human Behavior: Evolved Dispositions versus Social Roles.” *American Psychologist* 54 (6). American Psychological Association Inc.: 408–23

Ebenstein, Avraham. 2014. “Patrilocality and Missing Women.” *SSRN Electronic Journal*

Esteve-Volart, Berta. 2004. “Gender Discrimination and Growth: Theory and Evidence from India.” *LSE STICERD Research Paper*, no. DEDPS 42: 1–53.

Evans, Phil. 1999. “Occupational Downgrading and Upgrading in Britain.” *Economica* 66 (261):

79–96.

Fagernäs, Sonja. 2010. “Labor Law, Judicial Efficiency, and Informal Employment in India.” *Journal of Empirical Legal Studies* 7 (2): 282–321.

Fang, Hanming, and Andrea Moro. 2011. *Theories of Statistical Discrimination and Affirmative Action: A Survey*. *Handbook of Social Economics*. Vol. 1. Elsevier B.V.

Fernandez, R., A. Fogli, and C. Olivetti. 2004. “Mothers and Sons: Preference Formation and Female Labor Force Dynamics.” *The Quarterly Journal of Economics* 119 (4): 1249–99.

Fields, Gary S. 2006. “The Many Facets of Economic Mobility.” In *Inequality, Poverty and Well-Being*, 123–42. London: Palgrave Macmillan UK.

Fields, Gary S. 2007. *Employment in Low-Income Countries: Beyond Labor Market Segmentation? Employment and Shared Growth: Rethinking the Role of Labor Mobility for Development*. In P. Paci & P. Serneels (Eds.) Employment and shared growth: Rethinking the role of labor mobility for development (pp. 23-36). Washington, DC: The World Bank.

Fields, Gary S., and Efe A. Ok. 1999. “Measuring Movement of Incomes.” *Economica* 66 (264): 455–71.

Fields, G. S. (1994). Labour institutions and economic development: A conceptual framework with reference to Asia [Electronic version]. In G. Rodgers (Eds.), *Workers, institutions and economic growth in Asia* (pp. 113-148). Geneva, Switzerland: International Labour Organization, International Institute of Labour Studies.

Figart, Deborah M. 2009. “Discrimination.” In *Handbook of Economics and Ethics*, 91–98. Edward Elgar Publishing.

Fletcher, Erin, Rohini Pande, and Charity Maria Troyer Moore. 2018. “Women and Work in India: Descriptive Evidence and a Review of Potential Policies.” *SSRN Electronic Journal*.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2011a. “Decomposition Methods in Economics.” In *Handbook of Labor Economics*, 4:1–102. Elsevier Inc.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2011b. “Decomposition Methods in Economics.” In *Handbook of Labor Economics*, 4:1–102.

Fortin, Nicole M. 2008. “The Gender Wage Gap among Young Adults in the United States: The

Importance of Money versus People." *Journal of Human Resources* 43 (4): 884–918.

Francavilla, Francesca, Gianna Claudia Giannelli, and Leonardo Grilli. 2013. "Mothers' Employment and Their Children's Schooling: A Joint Multilevel Analysis for India." *World Development* 41 (1): 183–95.

Gang, Ira N., Kunal Sen, and Myeong Su Yun. 2017. "Is Caste Destiny? Occupational Diversification among Dalits in Rural India." *European Journal of Development Research* 29 (2): 476–92.

Gille, Véronique. 2018. "Applying for Social Programs in India: Roles of Local Politics and Caste Networks in Affirmative Action." *Journal of Comparative Economics* 46 (2): 436–56.

Gindling, T H. 1991. "Labor Market Segmentation and the Determination of Wages in the Public, Private-Formal, and Informal Sectors in San José, Costa Rica." *Economic Development and Cultural Change* 39 (3): 585–605.

Goel, Manisha. 2017. "Inequality Between and Within Skill Groups: The Curious Case of India." *World Development* 93 (May): 153–76.

Goldin, Claudia. 1995. "The U-Shaped Female Labor Force Function in Economic Development and Economic History." In: *Investment in Women's Human Capital and Economic Development,* edited by T.P. Schultz. University of Chicago Press ; 1995. pp. 61-90.

Goldin, Claudia. 2014. "A Pollution Theory of Discrimination : Male and Female Differences in Occupations and Earnings." In *Human Capital in History: The American Record*, edited by Leah Platt Boustan, Carola Frydman, and Robert A. Margo, 313–48.

Government of India. 2006. "Social, Economic and Educational Status of the Muslim Community of India." New Delhi:Government of India

Guérin, Isabelle. 2013. "Bonded Labour, Agrarian Changes and Capitalism: Emerging Patterns in South India." *Journal of Agrarian Change* 13 (3): 405–23.

Guérin, Isabelle, G Venkatasubramanian, and S Kumar. 2015. "Debt Bondage and the Tricks of Capital." *Economic & Political Weekly EPW Published on Saturday* L (26 & 27): 11–18.

Guha-Khasnobis, Basudeb., and Ravi Kanbur. 2006. "Informal Labour Markets and Development." *Studies in Development Economics and Policy*. Palgrave Macmillan.

Günther, Isabel, and Andrey Launov. 2012. "Informal Employment in Developing Countries: Opportunity or Last Resort?" *Journal of Development Economics* 97 (1): 88–98.

Halim, Nafisa, Kathryn M. Yount, and Solveig Cunningham. 2016. "Do Scheduled Caste and Scheduled Tribe Women Legislators Mean Lower Gender-Caste Gaps in Primary Schooling in India?" *Social Science Research* 58 (July): 122–34.

Hankivsky, Olena. 2012. *An Intersectionality-Based Policy Analysis Framework*. Vancouver: BC: Institute for Intersectionality Research and Policy, Simon Fraser University.

Harris, Christine R., Michael Jenkins, and Dale Glaser. 2006. "Gender Differences in Risk Assessment : Why Do Women Take Fewer Risks than Men ?" *Judgment and Decision Making* 1 (1): 48–63.

Harris, John R., and Michael P. Todaro. 1970. "Migration, Unemployment and Development. A Two-Sector Analysis." *American Economic Review* 60 (1): 126–42.

Harriss-White, Barbara. 2003. *India Working: Essays on Society and Economy*. Contemporary South Asia. Cambridge: Cambridge University Press, 2002.

Harriss-White, Barbara. 2010. "Work and Wellbeing in Informal Economies: The Regulative Roles of Institutions of Identity and the State." *World Development* 38 (2): 170–83.

Hart, Keith. 1973. "Informal Income Opportunities and Urban Employment in Ghana." *The Journal of Modern African Studies* 11 (01): 61.

Hoff, Karla, and Priyanka Pandey. 2014. "Making up People—The Effect of Identity on Performance in a Modernizing Society." *Journal of Development Economics* 106 (January): 118–31.

Howard Frederick, Angela. 2010. "'Practicing Electoral Politics in the Cracks.'" *Gender & Society* 24 (4): 475–98.

Hussmans, R. 2004. *Measuring the informal economy: From employment in the informal sector to informal employment*, WP No. 53, Policy Integration Department, Bureau of Statistics, ILO, Geneva.

International Labour Organization. 1998. *ILO Declaration on Fundamental Principles and Rights at Work (Declaration)*. Available at: https://www.ilo.org/declaration/lang--en/index.htm

ILO. 2003. "Statistical Definition of Informal Employment: Guidelines Endorsed by the Seventeenth International Conference of Labour Statisticians (2003)." Geneva: ILO, 2010 Available at: http://ilo.org/public/english/bureau/stat/download/papers/def.pdf

ILO. 2010. *Women in Labour Markets: Measuring Progress and Identifying Challenges*. International Labour Office. Geneva: ILO, 2010

ILO. 2016a. "India Labour Market Update." *ILO Decent Work Team for South Asia and Country Office for India*. New Delhi: ILO.

ILO. 2016b. *Women at Work: Trends 2016.* International Labour Office – Geneva: ILO

ILO. 2016c. *World employment and social outlook: trends for youth*. *International labour organization*. International Labour Office Geneva: ILO

ILO. 2017a. "Child Labour in India- Fact Sheet." Geneva:ILO available at: https://www.ilo.org/newdelhi/whatwedo/publications/WCMS_557089/lang--en/index.htm

ILO. 2017b. *Women, Gender and Work. Vol 2*. Edited by Uma (editors); Lansky, Mark; Ghosh, Jayati; Méda, Dominique; Rani. Vol. 2. Geneva: ILO.

ILO. 2017c. *World Employment Social Outlook 2017*. Geneva: ILO

Ito, Takahiro. 2009. "Caste Discrimination and Transaction Costs in the Labor Market: Evidence from Rural North India." *Journal of Development Economics* 88 (2): 292–300.

Iyer, Lakshmi, Tarun Khanna, and Ashutosh Varshney. 2011. "Caste and Entrepreneurship in India." *Harvard Buiness School*. Vol. 12.

Jaffrelot, Christophe. 2009. "La Question Musulmane." *Projet* 310 (3): 43.

James J . Heckman. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153–61.

Jäntti, Markus, and Stephen P. Jenkins. 2015. "Income Mobility." In *Handbook of Income Distribution*, 1st ed., 2:807–935. Elsevier B.V.

Jayachandran, Seema. 2015. "The Roots of Gender Inequality in Developing Countries." *Annual Review of Economics* 7 (1): 63–88.

Jayachandran, Seema, and Ilyana Kuziemko. 2011. "Why Domothers Breastfeed Girls Less than

Boys? Evidence and Implications Forchild Health in India." *Quarterly Journal of Economics* 126 (3): 1485–1538.

Jensen, Robert. 2012. "Do Labor Market Opportunities Affect Young Women's Work and Family Decisions? Experimental Evidence from India." *Quarterly Journal of Economics* 127 (2): 753–92.

Kabeer, Naila. 2012. "Women's Economic Empowerment and Inclusive Growth : Labour Markets and Enterprise Development." *SIG Working paper 2012/1*.

Kalsi, Priti. 2013. "Seeing Is Believing – Can Increasing the Number of Female Leaders Reduce Sex Selection in Rural India?" *Journal of Development Economics* 126 (13): 1–52.

Kambhampati, Uma Sarada. 2009. "Child Schooling and Work Decisions in India: The Role of Household and Regional Gender Equity." *Feminist Economics* 15 (4): 77–112.

Kanbur, Ravi. 2009. "Conceptualising Informality: Regulation and Enforcement." *Indian Journal of Labour Economics* 52 (1): 33–42.

Karimullah, and U. Kalpagam. 2010. "Are Muslims Discriminated against in the Labour Market in India?" *Indian Journal of Labour Economics* 53 (1): 61–79.

Katz-Wise, Sabra L., and Janet S. Hyde. 2010. "Gender-Role Attitudes and Behavior Access the Transition to Parenthood." *Developmental Psychology* 46 (1): 18–28.

Kaul, Tara. 2018. "Intra-Household Allocation of Educational Expenses: Gender Discrimination and Investing in the Future." *World Development* 104: 336–43.

Kingdon, Geeta, and Jeemol Unni. 2001. "Education and Women's Labour Market Outcomes in India." *Education Economics* 9 (2).

Klasen, Stephan, and Janneke Pieters. 2015. "What Explains the Stagnation of Female Labor Force Participation in Urban India?" *The World Bank Economic Review* 29 (3): 449–78.

Klumpp, Tilman, and Xuejuan Su. 2013. "A Theory of Perceived Discrimination." *Economic Theory* 53 (1): 153–80.

Kulshreshtha, A. C. 2011. "Measuring The Unorganized Sector In India." *Review of Income and Wealth* 57 (SUPPL. 1).

Lahoti, Rahul, and Hema Swaminathan. 2016. "Economic Development and Women's Labor Force Participation in India." *Feminist Economics* 22 (2): 168–95.

Laiglesia, Juan R. de, and Johannes Jütting. 2009. *Is Informal Normal ? Is Informal Normal*. Development Centre Studies. OECD Publishing.

Lamont, Andrea E., Jeroen K. Vermunt, and M. Lee Van Horn. 2016. "Regression Mixture Models: Does Modeling the Covariance Between Independent Variables and Latent Classes Improve the Results?" *Multivariate Behavioral Research* 51 (1): 35–52.

Lancaster, Geoffrey, Pushkar Maitra, and Ranjan Ray. 2008. "Household Expenditure Patterns and Gender Bias: Evidence from Selected Indian States." *Oxford Development Studies* 36 (2): 133–57.

Lang, Kevin, and Jee-Yeon K Lehmann. 2012. "Racial Discrimination in the Labor Market: Theory and Empirics." *Journal of Economic Literature* 50 (4): 959–1006.

Lanjouw, Peter, Rinku Murgai, and Nicholas Stern. 2013. "Nonfarm Diversification, Poverty, Economic Mobility, and Income Inequality: A Case Study in Village India." *Agricultural Economics* 44 (4–5): 461–73.

Lee, Doris, Apo Leong, Rene Ofreneo, and Anoop Sukumaran. 2008. *Rights for Two-Thirds of Asia: Asian Labour Law Review 2008*. Asia Monitor Resource Centre 2008.

Leontaridi, Marianthi Rannia. 1998. "Segmented Labour Markets : Theory." *Journal of Economic Surveys* 12 (1): 63–101.

Lerche, Jens. 2012. *Labour Regulations and Labour Standards in India: Decent Work? Global Labour Journal*. Vol. 3.

Lewis, W. Arthur. 1954. "Economic Development with Unlimited Supplies of Labour." *The Manchester School* 22 (2): 139–91.

Long, Jason, and Joseph Ferrie. 2013. "Intergenerational Occupational Mobility in Great Britain and the United States since 1850." *The American Economic Review* 103 (4): 1109–37.

Luke, Nancy, and Kaivan Munshi. 2011. "Women as Agents of Change: Female Income and Mobility in India." *Journal of Development Economics* 94 (1): 1–17.

Madheswaran, S, and P Attewell. 2007. "Caste Discrimination in the Indian Urban Labour Market:

Evidence from the National Sample Survey.” *Economic and Political Weekly* 42 (1983): 4146–54.

Maertens, Annemie. 2013. “Social Norms and Aspirations: Age of Marriage and Education in Rural India.” *World Development* 47: 1–15.

Magnac Th. 1991. “Segmented or Competitive Labor Markets?” *Econometrica* 59 (1): 165–87.

Mahajan, Kanika, and Bharat Ramaswami. 2017. “Caste, Female Labor Supply, and the Gender Wage Gap in India: Boserup Revisited.” *Economic Development and Cultural Change* 65 (2): 339–78.

Maloney, William F. 2004. “Informality Revisited.” *World Development* 32 (7): 1159–78.

Mammen, Kristin, and Christina Paxson. 2000. “Women’s Work and Economic Development.” *Journal of Economic Perspectives* 14 (4): 141–64.

Kamala Marius, G Venkatasubramanian. “Exploring Urban Economic Resilience: The Case of A Leather Industrial Cluster in Tamil Nadu”. “ *Savoirs et Mondes Indiens Working Papers Series*

Mehrotra, Santosh. 2006. “What Ails the Educationally Backward States? The Challenges of Public Finance, Private Provision and Household Costs.” In *The Economics of Elementary Education in India. The Challenge of Public Finance, Private Provision and Household Costs*, edited by Santosh Mehrotra. India:Sage Publications.

Mehrotra, Santosh, Jajati Parida, Sharmistha Sinha, and Ankita Gandhi. 2014. “Explaining Employment Trends in the Indian Economy: 1993-94 to 2011-12.” *Economic & Political Weekly* 49 (32).

Menon, Nidhiya, and Yana Van Der Meulen Rodgers. 2009. “International Trade and the Gender Wage Gap: New Evidence from India’s Manufacturing Sector.” *World Development* 37 (5): 965–81

Miller, Barbara D. 1982. “Female Labor Participation and Female Seclusion in Rural India: A Regional View.” *Economic Development and Cultural Change* 30 (4): 777–94.

Mincer, Jacob. 1962. *Labor Force Participation of Married Women: A Study of Labor Supply*. *Aspects of Labor Economics*. Vol. 1..

Mincer, Jacob A. 1974. *Schooling, Experience, and Earnings*. National Bureau of Economic Research, Inc.

Mincer, Jacob, and Solomon Polachek. 1974. "Family Investments in Human Capital: Earnings of Women." *Journal of Political Economy* 82 (2, Part 2): S109–10.

Mitra, A., and S. Verick. 2013. *Youth Employment and Unemployment: An Indian Perspective*. International Labour Organization, DWT for South Asia and Country Office for India. - New Delhi: ILO, 2013

Mitra, Anirban, and Debraj Ray. 2014. "Implications of an Economic Theory of Conflict: Hindu-Muslim Violence in India." *Journal of Political Economy* 122 (4): 719–65.

Monsen, Erik, Prashanth Mahagaonkar, and Christian Dienes. 2012. "Entrepreneurship in India: The Question of Occupational Transition." *Small Business Economics* 39 (2): 359–82.

Morrill, Melinda Sandler, and Thayer Morrill. 2013. "Intergenerational Links in Female Labor Force Participation." *Labour Economics* 20: 38–47.

Moser, Caroline O. N. 1978. "Informal Sector or Petty Commodity Production: Dualism or Dependence in Urban Development?" *World Development* 6 (9–10): 1041–64.

Motiram, Sripad, and Nayantara Sarma. 2014. "Polarization, Inequality, and Growth: The Indian Experience." *Oxford Development Studies* 42 (3): 297–318.

Motiram, Sripad, and Ashish Singh. 2012. "How Close Does the Apple Fall to the Tree? Some Evidence from India on Intergenerational Occupational Mobility." *Economic and Political Weekly* 47 (40): 56–65.

Mukherjee, Dipa, and Rajarshi Majumder. 2011. "Occupational Pattern, Wage Rates and Earning Disparities in India: A Decomposition Analysis." *Indian Economic Review* 46 (1): 131–52.

Munshi, Kaivan, and Mark Rosenzweig. 2006. "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy." *American Economic Review* 96 (4): 1225–52.

Muthén, Bengt, and Tihomir Asparouhov. 2009. "Multilevel Regression Mixture Analysis." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172 (3): 639–57.

Neal, Derek A., and William R. Johnson. 1996. "The Role of Premarket Factors in Black-White

Wage Differences." *Journal of Political Economy* 104 (5): 869–95.

Neumark, David. 1988. "Employers' Discriminatory Behavior and the Estimation of Wage Discrimination." *The Journal of Human Resources* 23 (3): 279–95.

Nguyen, Huu Chi, Christophe J. Nordman, and François Roubaud. 2013. "Who Suffers the Penalty?: A Panel Data Analysis of Earnings Gaps in Vietnam." *Journal of Development Studies* 49 (12): 1694–1710.

Nichols, Austin. 2010. "Income Inequality, Volatility, and Mobility Risk in China and the US." *China Economic Review* 21 (SUPPL. 1): S3–11.

Ñopo, Hugo. 2004. "Matching as a Tool to Decompose Wage Gaps." *IZA Discussion Paper Series*.

Ñopo, Hugo. 2008. "Matching as a Tool to Decompose Wage Gaps." *Review of Economics and Statistics* 90 (2): 290–99.

Nordman, Christophe J, and François Roubaud. 2009. "Reassessing the Gender Wage Gap in Madagascar: Does Labor Force Attachment Really Matter?" *Economic Development and Cultural Change* 57 (4): 785–808.

North, Douglass C. (Douglass Cecil). 1990. *Institutions, Institutional Change, and Economic Performance*. Cambridge University Press.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14 (3): 693–709.

OECD. 2010. *Growth and Sustainability in Brazil, China, India, Indonesia and South Africa*. Edited by Luiz de Mello. OECD Publications.

OECD. 2017. *OECD Economic Surveys: India 2017*. OECD Economic Surveys: India. OECD.

Perry, Guillermo E, William F Maloney, Omar S Arias, Pablo Fajnzylber, and Andrew D Mason Jaime Saavedra-chanduvi. 2010. *Informality: Exit and Exclusion*. *World*. Vol. 57.

Petrin, Amil, and Kenneth Train. 2010. "A Control Function Approach to Endogeneity in Consumer Choice Models." *Journal of Marketing Research* 47 (1): 3–13.

Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *The American Economic Review* 62 (4): 659–61.

Ponthieux, Sophie, and Dominique Meurs. 2015. *Gender Inequality*. *Handbook of Income Distribution*. 1st ed. Vol. 2. Elsevier B.V.

Psacharopoulos, George, and Zafiris Tzannatos. 1989. “Female Labor Force Participation: An International Perspective” *The World Bank Research Observer* 4 (2): 187–201.

Rama, Martín, Tara Béteille, Yue Li, Pradeep K. Mitra, and John Lincoln Newman. 2014. *Addressing Inequality in South Asia*. World Bank Publications.

Ranganathan, Thiagu, Amarnath Tripathi, and Ghanshyam Pandey. 2016. “Income Mobility among Social Groups in Indian Rural Households: Findings from the Indian Human Development Survey.” *IEG Working Paper No. 368*, no. 368.

Joshi, Shareen, Nishtha Kochhar, and Vijayendra Rao. 2017. “Are Caste Categories Misleading?: The Relationship between Gender and Jati in Three Indian States.” *WIDER Working Paper 2017/132*. Helsinki: UNU-WIDER, 2017.

Reddy, A. Bheemeshwar. 2015. “Changes in Intergenerational Occupational Mobility in India: Evidence from National Sample Surveys, 1983-2012.” *World Development* 76 (December): 329–43.

Reich, Michael, David M. Gordon, and Richard C. Edwards. 1973. “Dual Labor Markets: A Theory of Labor Market Segmentation.” *American Economic Review* 63 (2): 359–65.

Reimers, Cordelia W. 1983. “Labor Market Discrimination Against Hispanic and Black Men.” *The Review of Economics and Statistics* 65 (4): 570.

Renaut, Alain. 2014. “Différences, inégalités, injustice. Une grille conceptuelle de la démocratie.” In *Inégalités et justice sociale*, 97–107. Recherches. Paris: La Découverte.

Rentería, E. , Souto, G. , Mejía‐Guevara, I. and Patxot, C. (2016), The Effect of Education on the Demographic Dividend. *Population and Development Review*, 42: 651-671.

Rothenberg, Alexander D., Arya Gaduh, Nicholas E. Burger, Charina Chazali, Indrasari Tjandraningsih, Rini Radikun, Cole Sutera, and Sarah Weilant. 2016. “Rethinking Indonesia’s Informal Sector.” *World Development* 80: 96–113.

Sahoo, Soham & Klasen, Stephan, 2018. "Gender Segregation in Education and Its Implications for Labour Market Outcomes: Evidence from India," *IZA Discussion Papers 11660*.

Salem, Mélika Ben, and Isabelle Bensidoun. 2012. “The Heterogeneity of Informal Employment and Segmentation in the Turkish Labour Market.” *Journal of the Asia Pacific Economy* 17 (4): 578–92.

Santosh, R. 2015. *Muslims in Contemporary India*. Edited by Knut A Jacobsen. *Routledge Handbook of Contemporary India*. Routldge H. Routledge.

Sarkar, Sandip, and Balwant Singh Mehta. 2010. “Income Inequality in India: Pre- and Post Reform Periods.” *Economic & Political Weekly* xlv (37): 45–55.

Sarkar, Sudipa & Sahoo, Soham & Klasen, Stephan, 2017. "Employment Transitions of Women in India: A Panel Analysis," *IZA Discussion P*apers 11086

Sen, Amartya. 2000. “Social Exclusion: Concept, Application and Scrutiny” *Social Development Papers No. 1* Office of Environment and Social Development. Asian Development Bank

Sengupta, Arjun, K.P Kannan, Ravi Srivastava, and V.K Malhota. 2009. *The Challenge of Employment in India: An Informal Economy Perspective*. National Commission for Enterprises in the Unorganised Sector: New Delhi

Sharma, Smriti. 2015. “Caste-Based Crimes and Economic Status: Evidence from India.” *Journal of Comparative Economics* 43 (1): 204–26.

Sorsa, Piritta. 2015. “Raising the Economic Participation of Women in India: A New Growth Engine?” *OECD Economics Department Working Papers 1185: 97–124*. Paris:OECD

Soto, Hernando De. 1989. *The Other Path : The Invisible Revolution in the Third World*. Harper & Row.

Stewart, Frances. 2016. “The Dynamics of Horizontal Inequalities.” *2016 Human Development Report Office Think Piece*, UNDP Human Development Report Office. New York:UNDP.

Sundaram, Aparna, and Reeve Vanneman. 2008. “Gender Differentials in Literacy in India: The Intriguing Relationship with Women’s Labor Force Participation.” *World Development* 36 (1): 128–43.

The World Bank. 2012. *Gender Equality and Development - World Development Report 2012*. The World Bank.

The World Bank. 2018. *India’s Growth Story, India Development Update*. The World Bank.

Thorat, Sukhdeo, and Paul Attewell. 2007. “The Legacy of Social Exclusion. A Correspondence Study of Job Discrimination in India.” *Economic and Political Weekly* 42 (41): 4141–45.

Thorat, Sukhdeo, and Katherine Newman. 2010. “Blocked by Caste: Economic Discrimination in Modern India.” *Oxford University Press*, 377.

Tomei, Manuela. 2003. “Discrimination and Equality at Work: A Review of the Concepts.” *International Labour Review* 142 (4): 401–18.

Töpfer, Marina. 2017. “Detailed RIF Decomposition with Selection - The Gender Pay Gap in Italy.” *Hohenheim Discussion Papers in Business, Economics and Social Sciences* 26-2017

UNDP, United Nations Development Programme. 2015. *Human Development Report 2015 Work for Human Development*.

Unni, Jeemol. 2016. “Skill Gaps and Employability: Higher Education in India.” *Journal of Development Policy and Practice* 1 (1): 18–34.

Unni, Jeemol, and Ravikaran Naik. 2013. “Measuring Informality of Employment in Urban India.” *The IndIan Journal of Labour Economics* 56 (4): 493–509.

M. Vijayabaskar, Padmini Swaminathan, S. Anandhi, and Gayatri Balagopal. "Human Development in Tamil Nadu: Examining Linkages." *Economic and Political Weekly* 39, no. 8 (2004): 797-802.

Wedel, Michel. 2002. “Concomitant Variables in Finite Mixture Models.” *Statistica Neerlandica* 56 (3): 362–75.

Woetzel, Jonathan, Anu Madgavkar, and Shishir Gupta. 2017. *India’s Labour Market, A New Emphasis on Gainful Employment*. McKinsey Global Institute

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Wooldridge, Jeffrey M. 2015. “Control Function Methods in Applied Econometrics.” *Journal of Human Resources* 50 (2): 420–45.

Wu, Kin Bing, Pete Goldschmidt, Christy Kim Boscardin, and Mehtabul Azam. 2007. “Girls in India: Poverty, Location, and Social Disparities.” In *Exclusion, Gender and Education: Case Studies from the Developing World*, edited by Maureen Lewis and Marlaine Lockheed, 119–43.

Yahmed, Sarra Ben. 2018. "Formal but Less Equal. Gender Wage Gaps in Formal and Informal Jobs in Urban Brazil." *World Development* 101 (January): 73–87.

Zacharias, Ajit, and Vamsi Vakulabharanam. 2011. "Caste Stratification and Wealth Inequality in India." *World Development* 39 (10): 1820–33.

List of Tables

List of Figures

Detailed Table of Contents