DISCUSSION PAPER SERIES

DP15778
(v. 2)

A Division of Laborers: Identity and Efficiency in India

Guilhem Cassan, Daniel Keniston and Tatjana Kleineberg

DEVELOPMENT ECONOMICS

MACROECONOMICS AND GROWTH

ISSN 0265-8003

A Division of Laborers: Identity and Efficiency in India

Guilhem Cassan, Daniel Keniston and Tatjana Kleineberg

Discussion Paper DP15778
First Published 07 February 2021
This Revision 14 February 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

A Division of Laborers: Identity and Efficiency in India

Abstract

Workers' social identity affects their occupation, and therefore the structure and prosperity of the aggregate economy. We estimate a general equilibrium Roy model of this phenomenon in the Indian caste system, where work and identity are particularly intertwined. New data on occupation, wages, and caste's traditional occupations and social status show that workers are over-represented in their traditional occupations and under-represented in socially higher or lower occupations. We consider counterfactuals removing castes' hierarchical and occupational links. Despite more efficient human capital allocation, aggregate output gains are small–in some counterfactuals negative–due to weaker caste networks and reduced learning across generations.

JEL Classification: E24, E71, J21, J62, O15

Keywords: N/A

Guilhem Cassan - guilhemcassan@gmail.com
University of Namur and CEPR

Daniel Keniston - dkeniston@lsu.edu
Louisiana State University

Tatjana Kleineberg - tkleineberg@worldbank.org
World Bank

Acknowledgements
The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. This work was supported by the Fonds Wetenschappelijk Onderzoek – Vlaanderen (FWO) and the Fonds de la Recherche Scientifique – FNRS under EOS project O020918F (EOS ID 30784531) and by CEPREMAP. We are grateful to Joseph Altonji, Mark Rosenzweig and Nicholas Ryan and audiences at NES, WEHC, LSU, UBC, UNamur, ULB, BREAD, Warwick, Chicago, USC, Tulane, SEA, ASSA, and STEG for helpful suggestions. We are extremely grateful to James Nye at the University of Chicago library for his invaluable help in accessing the 1911 Census data.

A Division of Laborers: Identity and Efficiency in India

Guilhem Cassan
University of Namur

Daniel Keniston
Louisiana State University

Tatjana Kleineberg
World Bank*

February 14, 2022

Abstract

Workers' social identity affects their occupation, and therefore the structure and prosperity of the aggregate economy. We estimate a general equilibrium Roy model of this phenomenon in the Indian caste system, where work and identity are particularly intertwined. New data on occupation, wages, and caste's traditional occupations and social status show that workers are over-represented in their traditional occupations and under-represented in socially higher or lower occupations. We consider counterfactuals removing castes' hierarchical and occupational links. Despite more efficient human capital allocation, aggregate output gains are small–in some counterfactuals negative–due to weaker caste networks and reduced learning across generations.

JEL Codes: E24,E71,J21,J62,O15

Keywords: Occupational Choice, Human Capital Allocation, Caste, Networks, Aggregate Productivity, Occupational Identity, India, Institutions

---

*The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. This work was supported by the Fonds Wetenschappelijk Onderzoek – Vlaanderen (FWO) and the Fonds de la Recherche Scientifique – FNRS under EOS project O020918F (EOS ID 30784531) and by CEPREMAP. We are grateful to Joseph Altonji, Mark Rosenzweig and Nicholas Ryan and audiences at NES, WEHC, LSU, UBC, UNamur, ULB, BREAD, Warwick, Chicago, USC, Tulane, SEA, ASSA, and STEG for helpful suggestions. We are extremely grateful to James Nye at the University of Chicago library for his invaluable help in accessing the 1911 Census data.

"...the Caste System is not merely a division of labour. It is also a division of labourers."

"... the Caste System [...] involves an attempt to appoint tasks to individuals in advance, selected not on the basis of trained original capacities, but on that of the social status of the parents."

The Annihilation of Caste, Ambedkar (1936)

1 Introduction

Work is more than a source of income: it is a part of identity and subject to social norms. Thus occupational choices are not purely economic, but rather the outcome of an individual's traits and socio-economic environment: ethnic background, personality, and social aspirations. The complexity of the occupational choice problem is widely recognized. However, the challenges of quantifying its non-economic factors have presented substantial barriers to estimating their importance. In this paper, we analyze occupational choices in the context of the Indian caste system. Each Indian caste is associated with a place in the social hierarchy and (usually) a single traditional occupation. This occupation was historically seen as the proper vocation for members of that caste in society–their "dharma". Over time, the prestige of a job became associated with the social status of its traditional workers, and a parallel occupational hierarchy developed. While the end of the link between caste and occupation has often been predicted (Srinivas, 2003), it remains salient in modern India. Traditional occupations, which are observable and exogenous from the perspective of any single individual, provide a unique opportunity to study the role of identity and hierarchy in occupational choice.

This study quantifies the importance of identity for an individual's occupational choice and the impact of these identity-influenced choices on the economy as a whole. The link between caste identity and occupational choice affects the aggregate economy via two distinct channels. First, the preference for one's traditional occupation as well as discrimination in alternative occupations can distort workers' occupational selection away from their comparative advantage, leading to an inefficient allocation of human capital. Second, the sorting of castes into traditional occupations can enable the transfer of occupation-specific skills from parents to children and the formation of social networks. Intergenerational learning and networks can lastingly affect workers' occupational choices and can create "path dependence" beyond workers' own preferences: even if workers no longer feel tied to their traditional occupations, they might nevertheless work in them to take advantage of the productivity effects of large caste networks and of working in the same occupation as their fathers. Unlike the distortionary effects on the allocation of human capital, the aggregate impact of these channels of historical persistence can be positive and essential to explaining the remarkable endurance of occupational identity and hierarchical divisions over time.

We first document a series of new empirical facts that illustrate the role of caste membership for occupational choices and wages. We find that individuals are about three times as likely to work in their traditional occupation compared to any other occupation. These effects grow stronger as the hierarchical difference between the prestige of the alternative occupation and the social status of the caste increases. Within a caste, workers employed in their traditional occupation earn less than their caste-mates who work in other occupations. However, when we examine earnings within occupation (i.e., controlling for occupation fixed effects) we find that workers employed in their traditional occupation earn more than workers from other castes who work in the same occupation. The data further show that returns to ability–measured by schooling and experience–are lower in common traditional caste occupations compared to "modern" occupations. These empirical findings are informative about how workers select into occupations based on their comparative and absolute advantage, which we formalize in our model.

We develop and estimate a structural general equilibrium Roy (1951) model of education and occupational choice that incorporates caste identity through several channels: a direct preference for traditional occupations and against lower status occupations, discrimination in higher status occupations, productivity effects from working in one's father's occupation, and network effects at the caste-occupation level. As in the standard Roy model, workers differ in productivity which varies independently across occupations so that workers select into occupations based on their comparative advantage. We extend the model by allowing workers to further differ in general ability. Individuals with high general ability should work in "modern" high-return occupations; however, caste identity can draw them back into their traditional occupation in which we find low returns to ability. Allowing selection on both occupation-specific and general ability is essential to accommodate our empirical findings in the Roy model.

To study the importance of castes' occupational links and to quantify their aggregate effects, we consider the economy in a general equilibrium context. Since wages reflect the marginal product of human capital in an occupation, it is essential to allow wages to adjust when estimating the effects of reallocating human capital across occupations. In addition, occupational choices are closely linked to individuals' education choices and to the composition of social networks. We therefore determine wages, educational choices, and social networks endogenously. The general equilibrium nature of our analysis allows for the possibility that castes' occupational identity can serve as a means of equilibrium selection, that can help workers to coordinate on a human capital allocation and a network composition that maximizes output (Chen and Chen, 2011).

To estimate the model, we construct a novel data set that links contemporaneous micro-data on occupational choices, wages, and demographics to detailed historical data on castes' traditional occupation and hierarchical status. We then use our estimated model to investigate the importance of two dimensions of caste identity: *occupational identity*, which links castes to their traditional occupation, and *social status*, which entails discrimination against workers in occupations that rank socially higher than their caste, and workers' distaste for occupations that rank socially lower than their caste. First, we remove direct attachment to traditional occupations, or the direct effects of castes' social status (i.e., workers' discrimination in lower-ranked occupations and their disutility in higher-ranked occupations). Second, we remove effects through parental occupations and intergenerational learning. Last, we allow caste-occupation networks to adjust endogenously.

Removing castes' direct attachment to traditional occupations has very small aggregate effects when holding parental occupations and caste-occupation networks constant: aggregate output increases by only 0.3 percent. Removing the direct effects of caste hierarchy has larger effects, increasing output by 3.8 percent, and removing both aspects simultaneously leads to the largest change, increasing output by 4.1 percent. Aggregate effects are relatively small because workers who leave their traditional occupations and occupations linked to their castes' hierarchical position lose the larger caste networks and intergenerational knowledge transfers found in these occupations.

We next eliminate castes' occupational attachment due to intergenerational learning. Removing the correlation between fathers' occupations and traditional occupations decreases output by 7%, while removing the correlation between fathers' occupations and occupations' social hierarchy causes a 2.5 percent decrease.Removing both lowers output by 3.4 percent. Effects are negative because workers no longer have the option of simultaneously choosing their father's occupation and the occupation where their caste networks are largest.

Finally, we remove the last channel of castes' occupational attachment by adjusting caste-occupation networks endogenously. When removing links to traditional occupations, losses in aggregate output are now even larger with a drop of 9.8 percent; however, when removing caste hierarchy, the adjustment of networks dampens the negative effect on aggregate output from -2.5 percent with exogenous networks to -1.9 percent with endogenous networks. This demonstrates the importance of traditional occupations in coordinating castes into one single occupation for network formation. When removing both preferences for traditional occupation and caste hierarchy, output drops by 7.1 percent. Ultimately, productivity losses from weaker networks and less intergenerational learning dominate gains from an improved selection of workers based on their comparative advantage and reduced bias.

The paper proceeds as follows. Section 2 describes the anthropology and origins of the caste system and reviews the relevant literature. Sections 3 and 4 describe the data and our reduced form analysis. Section 5 presents our model. Section 6 describes the estimation strategy. Section 7 presents the estimation and counterfactual results and Section 8 concludes.

2 Caste and the Labor Market

The Indian caste system: Origins and Anthropology

Srinivas (1962) defines a caste, or *jati*,[1] as "a hereditary, endogamous, usually localized group, having a traditional association with an occupation, and a particular position in the local hierarchy of castes." This definition highlights two aspects of caste: first, it is an ordering of social prestige which associates castes with the principles of pollution and purity (Dumont, 1970); second, it is a division of labor across occupations which became hereditary and endogamous (Ibbetson, 1916; Gupta, 2000; Bidner and Eswaran, 2015; Beteille, 1996).[2]

The hierarchical nature of caste emerges in the earliest Hindu texts. A 200 BCE compilation of

[1] *Jatis* are narrowly defined caste groups which are the relevant dimension of caste identity for most Indians (Vaid, 2014).

[2] A related literature, initiated by Wiser (1936), studies the "jajmani" system which highlights the patterns of caste based patron-client relationships and occupational specialization.

religious and social laws, the Manusmriti, defines a broad and inflexible hierarchy with Brahmans at the top and Shudras at the bottom.[3] Work by modern anthropologists (Srinivas, 1994; Deliege, 1993; Michelutti, 2008) as well as contemporaneous surveys confirm the continued importance of caste hierarchy: in the 2011/12 India Human Development Survey (IHDS), 43 percent of the members of the ritually highest Brahman caste report that someone in their household practices "untouchability", defined as avoiding all contact with "ritually polluting" lower castes. More generally, the caste hierarchy continues to shape social interactions in modern India, including the language used in conversations, the sharing of food, and marriage practices. In the post-independence era, the Indian government acknowledged the historical oppression of certain castes. To remedy this discrimination, the government instituted affirmative action programs for three "backward" groups which were categorized as scheduled castes (SCs), scheduled tribes (STs), or Other Backward Castes (OBCs).

The association of castes with a traditional occupation has an equally long history, with its virtues acclaimed by Krishna in the central Hindu text of the Bhagavad Gita. Proponents of the caste system, most famously Mohandas Gandhi and Swami Vivekananda, argued that the true nature of caste was occupational, which could be separated from its oppressive hierarchical aspect. "But if varna [caste] reveals the law of one's being and thus the duty one has to perform, it confers no right, and the idea of superiority or inferiority is wholly repugnant to it." (*Harijan,* 1934, cited in mkgandhi.org). Academics have extensively documented the importance of traditional occupations.[4] A consistent theme in this work is the emotional link that caste members feel with their traditional occupation, as described by the words of a Labbai mat weaver in Tamil Nadu: *"indha thozhil enga rathathil oori irukku* (roughly, this profession is present in our blood)" (Venkatesan, 2006, p. 73). Castes' traditional occupations are common knowledge and remain salient even for caste members who work in other occupations (Deliege, 2004). For example, Doron (2013) writes "there is certainly a sense among Mallahs of all occupational designations in Banaras that plying boats represents a kind of archetypical or traditional Mallah occupation" (p. 90). Anthropological studies have documented that the development of novel economic opportunities has changed the associations between castes and occupations, maybe loosening but certainly not eliminating them (Ghurye, 1961; Mayer, 1996; Bayly, 1999; Beteille, 2012).[5]

Critics of the caste system (Ambedkar, 1936) emphasize that the hierarchical and occupational nature of caste are deeply intertwined. Most high wage occupations were historically forbidden to lower castes, while occupations associated with lower castes tend to be unpleasant, servile, and offer low returns (e.g., waste removal). Other occupations, for example cooking, have become linked with

[3]In the subsequent millennia, castes outside of the traditional Hindu rankings, *Panchamas,* came to occupy the lowest place in the hierarchy.

[4]This literature includes economic studies of villages that span multiple decades (Mayer, 1996; Lanjouw and Stern, 1998) and a wide array of anthropological case studies. Recent examples include work on Mallah caste boatmen (Doron, 2013), Ossan caste barbers (Amir, 2019), and Kumar caste potters (Heierstad, 2017).

[5]Beteille (2012) writes "No handicraft could sustain the entire population of the caste which was associated with it, particularly when the population was rising and where opportunities for migration were limited. The surplus population from a particular caste or sub-caste could always move into agriculture or some other gainful activity not associated with any particular caste. What was not easy was the movement from one to another specialised craft or service already assigned to an existing caste or sub-caste."

upper castes due to their inherent association with ideals of purity (Iversen and Raghavendra, 2006). Conversely, occupations associated with low castes, such as leather work or goods transportation, may become stigmatized for high castes. Aggarwal et al. (2015) document in a case study that elite positions in business, press, and academia are dominated by upper castes, while jobs involving manual labor are exclusively done by lower castes. The hierarchy of castes has thus given birth to a hierarchy of occupations, with certain occupations being considered pure and desirable, and others polluting and degrading (Mayer, 1996; Mosse, 2020).

The concept of caste in South Asia originates in Hinduism, but has extended across other religions. Non-Hindu castes are similarly characterized by a hierarchical order and an occupational association with examples including Muslim weavers (Ansari), Christian fishermen (Paravas), and others. Muslim castes are often broader than Hindu castes, but they retain the same defining concept of endogamy. Low caste Sikhs and Buddhists benefit from the same government affirmative action policies as Hindu scheduled castes, and Dalit Christians have long sought these benefits.

Economics literature on identity and the Indian caste system

The concept of occupational identity, and its link with social hierarchy, are the building blocks of the anthropological literature cited above and are widely studied in Western sociology and psychology (see Skorikov and Vondracek (2011) for a recent review). In economics, empirical evidence on the effects of identity is less developed. However, theoretical work has studied the role of identity and social norms on economic behavior (e.g., Akerlof and Kranton (2000); Akerlof (1976, 1980)). Akerlof (1980) shows that the fear of losing reputation can prevent individuals from making economically optimal labor market choices if such choices imply deviating from social norms. In line with our findings, Akerlof (1976) points out that removing a taste for following widely shared social norms may not be enough to alter behavior due to the fear of social sanctions.

The economic literature on the Indian caste system focuses on its effects on risk sharing, social networks, and intergenerational skill transmission. Munshi (2019) provides a survey of this literature. Munshi and Rosenzweig (2006) examine the educational choices of Mumbai residents, arguing that lower castes discourage their most able young men from pursuing high skill occupations to preserve strong social networks in low skill traditional occupations. Banerjee and Munshi (2004a) provide a case study of the knitted garment industry in Tirupur and show that community ties determine firms' capital investments. Several studies show that intergenerational transmission is remarkably strong for education (Borkotoky et al., 2015) and occupation (Kumar et al., 2002; Deshpande and Palshikar, 2008; Vaid, 2012; Hnatkovska et al., 2013; Iversen et al., 2017). Kumar et al. (2002) and Vaid (2012) document that castes play an important role in the intergenerational transmission of occupation, which does not seem to weaken over time. Oh (2021) studies the link between jatis and their traditional occupations with an experiment. She finds that workers are significantly less likely to accept casual labor tasks which are not linked to their traditional occupation, especially if these tasks are associated with a hierarchically inferior caste. Our paper complements this micro evidence by analyzing the macro-economic effects of the Indian caste system on human capital allocation and aggregate output, while taking into account how occupational wages, schooling decisions, and caste networks adjust in general equilibrium.

Economics literature on the allocation of human capital

Our work further relates to studies that explore the aggregate implications of frictions to human capital allocation. Hsieh et al. (2019) might be closest to our paper in spirit, as they quantify the effects of decreased discrimination against women and Blacks in high skill-return occupations on aggregate wage and GDP growth in the United States. It is common in this literature to assume that occupation-specific productivity is uncorrelated across occupations. Under this assumption, expanding sectors increasingly attract workers with lower comparative and absolute advantage, while contracting sectors shed the least productive workers (as noted by Young in his 2014 study of US industries). Alvarez-Cuadrado et al. (2020) re-examine the selection into agriculture at the micro level, and find that the most productive agricultural workers are also more productive in their secondary occupations.[6] We add to this literature by empirically testing the model implications of different sorting mechanisms with data on occupational choices and wage distributions across and within castes. One contribution of our paper is to combine a general equilibrium approach with a structural maximum likelihood estimation that uses rich individual-level data.

3 Data

In the Indian context, workers' identity and social networks are defined by narrowly defined caste groups, *"jatis"*, rather than by the larger *varna* caste groupings or the government reservation categories. The Indian Household Development Survey (IHDS) is one of the few data sets which provides information on jati in addition to individuals' demographics, occupations, wages, and family characteristics. We complement this data set with three sources. First, we construct social networks at the jati-occupation level with data from the Demographic and Health Survey (DHS, also called NFHS) (IIPS, 2007). Second, we retrieve information on each jati's traditional occupation from the colonial Census in 1911 which we complement with other historical sources. Third, we use a unique 1901 Census ordering of castes to establish a hierarchical ranking of castes and occupations. Merging these four data sets at the jati level poses particular challenges and requires a very labor-intensive harmonization of jati names. In the following paragraphs, we describe the construction of our data set and our strategy of merging jati names across different data sources in more detail.

Harmonization of Jati names

The IHDS and DHS report jati names declared by respondents verbatim. This complicates classification because the meaning of "caste" itself can be ambiguous (Headley, 2013), and because

[6] In spatial general equilibrium, Eckert and Peters (2018), Heise and Porzio (2021) and Bryan and Morten (2019) find that the Roy model with uncorrelated shocks can not explain the patterns of regional migration and productivity in the data.

there are many synonyms and spellings for each jati, not to mention typos. To categorize jati names in a systematic manner, we use the People of India project which was launched in 1985 by the Anthropological Survey of India and which made an extraordinary effort to systematically collect data on all Indian jatis. The project produced a volume (Singh, 1996) that lists 2,205 "Main Communities" (which correspond to a *jati*) and their various synonyms at the state level. We digitized this volume to create a jati "master list" with state-specific lists of all jati synonyms. We then hand-merged this master list with the IHDS and DHS, with the help of several research assistants, ultimately categorizing 32,137 recorded names into the 1,167 unique castes which we use in the main analysis of the IHDS data.[7]

Individual-level data from the IHDS Household Survey

This project's primary data set is the 2011 round of the Indian Household Development Survey (IHDS) (Desai et al., 2008; Desai and Vanneman, 2015). The survey provides rich demographic data on 42,152 households. The extensive occupation and income module records income and time spent in each occupation for each individual in the household.[8] The survey also documents the occupation of the household head's father or, in female headed households, father-in-law. When this occupation is missing, we impute it using a multiple imputation technique.[9] We add information on jatis' social status by matching jati names to their current classification as Scheduled Castes (SC), Scheduled Tribes (ST) or Other Backward Classes (OBC), which are rough proxies of social ranking.[10] To do so, we follow Cassan (2019) and use the classifications from states' official reservation lists.[11]

Occupation-specific caste networks from the DHS Household Survey

The estimation of social networks at the jati-occupation-level for over 1,000 jatis and 49 occupations requires a very large sample size. We therefore construct networks by combining the IHDS data with the third round of the DHS (2005-06) which provides jati and occupation information for the

[7]The discrepancy between the total number of POI Main Communities (2,205) and those found in our sample (1,167) is due to the fact that not all communities are reported by IHDS respondents. The POI also enumerates many tribal groups from India's Northeastern states which are not included in our analysis. We were unable to link 7% of IHDS individuals to a caste. These fall into 2 categories. 1% reported a caste not identifiable in the IHDS–we drop these observations. 6% reported a non-caste category when asked about caste, most frequently the name of their religion (e.g. "Muslim"), but occasionally the name of a region (e.g. "Punjabi"). We retain these observations but do not link them to a traditional occupation, nor do we construct their caste-based social networks.

[8]The IHDS is a panel data set with two rounds, where we use the first round (2005-06) only to complement missing or incomplete data in the second round. For example, we use information on jati names or parental occupation from the first round if these variables are missing (or imprecise) in the second round. Income and time use data on secondary occupations, home work, animal care, money lending, and land rental posed specific challenges in the cleaning and construction of our final data set, which we explain in Appendix A1.

[9]Information on father's occupation is missing for 12.8 percent of men and 84.7 percent of women. In Appendix A2.4, we use data from a different survey, which records parental occupations for all respondents, to show that our main results are not affected by this imputation.

[10]These groups are eligible for affirmative action policies by the Indian government. SCs were historically most discriminated, STs are aboriginal tribes with limited access to public goods, OBCs are low in the caste hierarchy but were subject to less discrimination than SCs.

[11]This approach ensures that there is no variation of the reservation status within jati-state.

(female) respondent and her spouse in a sample of 109,041 households. We use the DHS only for the social network data because it contains very sparse information on income and parental occupation.

Traditional occupations from historical data sources

We obtain information on jatis' traditional occupations in each Province (the colonial equivalent of a state) from the colonial Census of 1911 (Conlon, 1981).[12] We complement the data with several other historical data sources to improve the completeness of the data set, primarily using Kitts (1885), which is based on the 1881 Census.[13] Our linking of jatis to their traditional occupations is largely consistent between the 1811 and 1911 data sources: 79 percent of the jatis that are present in both lists (representing 88 percent of the 1911 population) are matched to the same traditional occupation, despite differences between data sources in occupational categories, anthropological approaches, and geographical classification of castes. The most common differences in traditional occupations can be explained by the colonial policy of sedentarizing tribal and nomadic communities (Bayly, 1999): 34 groups in 1881 are listed as "forest and hill men", out of which 26 are reclassified as "agriculturalists" or "agricultural labourers" in 1911.[14]

Census ethnographic accounts confirm that traditional occupations were assigned on the basis of reported caste identity, rather than castes' actual occupation. For example, the Kewat caste in the United Provinces is classified as "Boatmen, fishermen and riverain occupations", although less than 1 percent of its male members actually worked in that occupation. Even if the classification of traditional occupations had been influenced by castes' actual employment in 1911, their relevance today–100 years later–could still be considered the result of tradition and path dependency. Similarly, the modern IHDS data show some castes have re-coordinated into non-traditional occupations. For example, almost 1/3rd of the Kulala caste (traditionally potters) now work in the tobacco industry, a pattern also documented in the anthropological literature (Shankar and Singhe, 2014). We treat this phenomenon as a social network, rather than a traditional occupation. Our analysis thus focuses on the long-run misallocation effects of caste identity as observed in 1911. While we do not estimate the full effect of individuals' contemporaneous identity, we capture those aspects of it which are most resistant to change and likely to create the most frictions in long-run development (Kranton, 2016).

We create a crosswalk between historical and modern (NCO68) occupation codes by defining 49 consistent occupational categories (listed in Appendix Table A9).[15] These categories are frequently broader than the Census definitions of traditional occupations. For example, a caste whose traditional

[12]We use the tables titled "Occupation by selected castes, tribes or races". For the Provinces of Rajputana and Ajmer-Merwara, we use the Census of 1921 instead, as the 1911 Census did not offer a good coverage of jati's traditional occupations.

[13]If jatis are missing in both data sets, we use the People of India "India's Communities" volume which provides rich historical and anthropological information about all jatis–usually including jatis' traditional occupation.

[14]For jatis whose traditional occupation was labeled as "criminal" in colonial era sources, we found historical evidence that these groups were nomadic and subject to state-level discrimination that aimed at sedentarizing those groups (Schwarz, 2010). For these jatis, we instead retrieve their traditional occupation from data in Crooke (1896) or–if missing there–from the People of India volume.

[15]In the structural estimation, we drop the "Beggars" occupational category, which has only 4 individuals in our sample.

occupation was sweet making is matched to the category of "food and beverage production". This binning may lead us to understate the importance of traditional occupations if individuals feel an identity link only to their traditional occupation but not to the other occupations in the broader occupation category. Intuitively we might estimate a moderately increased probability that castes select the broad occupational category (which contains their traditional occupation) rather than a greatly increased probability that castes select their exact narrow traditional occupation.

Caste and occupational rankings from 1901 Census

The 1901 Census of India provides a unique source of hierarchical rankings of castes. Census officials, assisted by Indian elites, grouped castes into five to twelve ordered hierarchical categories within each province of British India. Occupation is listed as one (out of eleven) potential bases of this ranking, but the hierarchy is clearly centered upon the degree to which castes are viewed as ritually polluting by the priestly Brahman caste.[16] Muslims are also ranked into three to five groups, according to traditional categories of *ashraf*, *ajlaf*, etc. The ranking was controversial at the time, provoking a flood of petitions and objections from caste associations who felt their caste rankings did not reflect the true worth (or aspirations) of their community (Lee, 2019). Some have argued that this enumeration and ranking of castes served to rigidify the caste system which had previously been much more flexible (Cohn, 1987; Dirks, 2001). Despite some shortcomings, the 1901 rankings provide a much more detailed measure of castes' social hierarchy than the government's modern reservation categories, and they are not influenced by social and political activism from caste associations themselves.[17]

We use these historical caste rankings to develop corresponding occupational rankings. For each province and occupation, we identify all local castes for which that occupation was traditional. We then compute the average social ranking of those castes (weighted by caste population within province) and assign it to the occupation. Thus occupations for which the traditional workers are low status receive a low occupational ranking (e.g., leatherwork), while occupations associated with workers high in the hierarchy receive a high rank (e.g., teaching).[18] Occupations that have no traditional workers remain unranked (e.g., engineers, or tobacco workers). Appendix Table A9 displays the national average of these province-level rankings. Consistent with the anthropological literature, we find that education and religious occupations rank highest, while leathermaking and sanitation work rank lowest. Appendix A1 contains additional details on the assembly of the

[16] For instance, in North India a middle rank category is titled "Castes from whom some of the twice-born take water and *pakki* [cooked food], without question", while in South India a lower caste is labeled "Sudras who do not employ Brahman purohits and whose touch pollutes".

[17] There is ample evidence that castes and tribes attempted to become included in the government's current reservation lists to benefit from affirmative action programs. One example is the Maratha caste which was included in 2019 in the list of "Other Backward Castes" (OBC) in the state of Maharastra after years of lobbying and violent agitation. In 2021, this ruling was overturned by the Indian Supreme Court on the grounds that "The Marathas are dominant forward class and are in the main stream of National life." (*Jaishri Laxmanrao Patil vs The Chief Minister And Ors.* 5 May, 2021).

[18] In this ranking, we aim to capture the observations of sociologists such as Ghurye (1961), who writes "Almost universally it was the group of non-polluting occupations commonly believed to be the open field for the non-polluting upper castes round which popular valuation was focused." (p. 242)

hierarchical data.

4 Reduced Form Evidence

Traditional occupation and occupational choice

We begin our empirical analysis by documenting the extent to which the traditional occupation and social position of a jati determines the contemporary occupational choice of its members. Figures 1a and 1b show the share of men and women who work in their jati's traditional occupation and the share who works in their father's occupation. The figures show that "traditional workers" are over-represented in their traditional occupations, often by a large degree. One such occupation is "dyeing and cleaning" in which 60 percent of male and 80 percent of female workers follow their traditional occupation. To put this into perspective, the figures further compare the observed shares of traditional caste workers in each occupation to the shares of these castes that a random allocation would imply (holding the existing occupational structure constant). For many occupations, observed shares are multiple times larger, e.g., up to 40 (60) times larger for male (female) traditional workers in "dyeing and cleaning". Other occupations have close to no traditional workers–such as legal or medical professions. Overall, 17.3 percent of men and 8.1 percent of women work in their jati's traditional occupation compared to 9.5 percent of men and 5.1 percent of women under a random allocation.[19] The concentration of castes in their traditional occupation was even higher a century ago (in 1911) with 48 percent of men and 37 percent of employed women working in their traditional occupation.[20]

Among non-Hindus men, 12.4% work in their traditional occupation, almost double the 6.6% that would occur under random allocation; for women shares are much smaller at 2.4% and 1.9%, respectively. Over-representation in the traditional occupation is thus not specific to Hindus only, as suggested in the literature.

Workers are also over-represented in their father's occupation, however, this effect is weaker for women. Occupations with a higher share of traditional workers tend to have more workers who follow their father's occupation, but these outcomes are not perfectly correlated. In occupations with strong caste links, the share of traditional workers is often larger than the share in their parents' occupation. Thus 33 percent of religious workers and 23 percent of barbers are traditional caste workers whose parents were employed in different, non-traditional occupations. Traditional occupations seems to retain their appeal, even among members of the current generation whose parents worked in other occupations.

These findings suggest that caste identity is closely linked to traditional occupations. An alternative explanation could be that consumers prefer to buy products from traditional castes, whose "vocation" it is to produce that good. This may be true for some cases (e.g., religious workers),

[19]These ties are stronger in rural areas with 20.2 percent of rural men and 11.2 percent of rural women in their traditional occupations.

[20]Conditioning on the same sample of jatis that is available in the 1911 census and the contemporary IHDS, this compares to 18 percent of men and 10 percent of women today.

however, we also see large shares of traditional workers in occupations where the producer is unknown to consumers (e.g., fishing, jewelry, cultivation). Oh (2021) confirms in an experimental study that workers' preference for their traditional occupation and disinclination to work in a socially inferior occupations is not affected by whether occupational choices are made in public or private.

To quantify the effect of traditional occupational preferences more formally, we turn to a regression analysis. We rectangularize our data set at the individual×occupation level, so that each individual is observed once for each potential occupation (i.e., 49 times). We then run the following OLS regression:

$$Occ_{iok} = \alpha + \beta TradOcc_{ok} + \pi OccAboveCaste_{ok} + \lambda OccBelowCaste + \gamma_o + \varepsilon_{iok},$$

where Occ_{iok} indicates that individual i of jati k works in occupation o, $TradOcc_{ok}$ indicates that occupation o is jati k's traditional occupation, and $OccAboveCaste_{ok}$ and $OccBelowCaste_{ok}$ measure the (absolute) difference between the social ranking score of occupation o and the social ranking score of respondent's jati k. $OccAboveCaste_{ok}$ is positive if an occupation ranks higher than the individual's caste and zero otherwise. $OccBelowCaste_{ok}$ is positive if the occupation is socially inferior to the respondent and zero otherwise. If an occupation is not associated with any historical status level, both variables are zero. γ_o are occupation fixed effects. We bootstrap standard errors, clustering the bootstrap at the PSU level in accordance with the 2-stage sampling of the IHDS, following Abadie et al. (2017). We recalculate the social network variable with each bootstrap iteration to account for variation in this generated regressor. We integrate multiple imputations with bootstrapping following the "Boot MI" technique in Schomaker and Heumann (2018).

Table 1 presents the results. Column 1 of Panel A shows that men are 6.7 percentage points more likely to work in an occupation if it is their jati's traditional occupation, holding constant occupational and individual characteristics. In Columns 2-5 we include measures of the absolute difference between the social status of an occupation and the individual's caste's social position. Relative to unranked "modern" occupations, male workers are less likely to work in occupations that are either below or above their caste in social status. However, these effects are small: estimated coefficients range from -0.009 to -0.016 and social rankings are scaled (0,1), so the largest possible effect would reduce the choice probability by just 1.6 percent. Columns 3-5 further include individuals' jati network, defined as the share of workers in their occupation who belong to their caste, and an indicator for whether they work in the same occupation as their father. We find that both variables are significant and large in magnitude: the probability that male workers choose an occupation increases by 31 percentage points if it is their father's occupation; and by 7-8 percentage points for each 10 percentage point increase in the occupation's caste network (Column 4). The impact of traditional occupations remains significant with these controls but decreases to roughly 4 percentage points. This finding implies that workers are three times more likely to work in their traditional occupation compared to a random occupational choice. Finally, column 5 examines heterogeneity. Male workers of scheduled castes (SCs) have less affinity for their traditional occupations, while the appeal of the traditional occupation is much stronger (12 percentage points) for workers whose fathers also work in the traditional occupation. Parents may transmit not just skills, but also traditional values, and once the older generation has deviated from traditional occupation it loses some (though not all) of its appeal for their children.[21]

Panel B of Table 1 shows that effects are present but much smaller for women. On average, women are 2.6 percentage points more likely to work in their traditional occupation, which reduces to 1 percentage point when controlling for father's occupation and caste-occupation networks. Including occupational rankings in columns 2-5 suggests that women are less likely to work in occupations that rank lower than their caste, but equally likely to work in occupations that rank higher than their castes, after controlling for social networks. Women's traditional occupation choice probability increases by 11 percentage points for their father's occupation and by 4 percentage points for each 10 percentage point increase in the occupation's caste network (Column 4).

We perform several robustness checks. In Appendix Tables A2-A5, we re-estimate our main regressions with different network measures. First, we compute a network measure that excludes observations from each individual's current state of residency. This procedure addresses the possible concern that a caste's network within an occupation may be correlated with local factors that can affect the caste's presence (and productivity) in the occupation. One example could be distance from the coast for a fishing caste. Second, we compute networks based on father's occupations which may better reflect caste's exogenous social networks. In Appendix A2.2 and Table A1, we further test the possibility that family members other than one's father can affect occupational choices through the transmission of skills or physical capital. We find that these variables are often significant, but they do not change the estimated effects of traditional occupations or social status on individuals' occupational choices.

Selection and productivity in traditional occupations

To examine the relationship between occupational identity and hourly wages, we run the following regression:

$$\log(wage/hour)_{iok} = \alpha + \beta TradOcc_{iok} + \lambda OccBelowCaste_{ok} + \gamma X_{iok} + \varepsilon_{iok}$$

where $\log(wage/hour)_{iok}$ is the log of hourly wages of individual i from jati k working in occupation o, $TradOcc_{iok}$ indicates whether occupation o is the traditional occupation of worker i's jati, $OccBelowCaste_{ok}$ quantifies the extent to which occupation o was historically considered to be socially inferior to worker i's jati, and X_{iok} is a set of individual characteristics which include father's occupation and caste networks. Table 2 presents the results for two specifications, controlling first for jati fixed effects and then for occupation fixed effects. All specifications condition on labor force participation, which particularly affects sample size for women.

With jati fixed effects (Columns 1), we find that men's hourly wage is 21 percent lower in their traditional occupation, compared to workers from the same jati who work in any other (non-

[21]We thank an anonymous referee for this insight.

traditional) occupation. The result holds for women with a slightly smaller coefficient (Column 3). This finding is consistent with the standard selection effects of the uncorrelated Roy model in which the first workers who enter an occupation have the highest occupation-specific productivity. Average occupation-specific productivity of a caste would therefore decrease if preferences drew more of its workers into a given occupation.[22] Similarly, the coefficient on $OccBelowCaste_{ok}$ show that caste members who work in socially inferior occupations are compensated with higher wages–much higher in the case of women.

Perhaps surprisingly, the results are reversed with occupation fixed effects (Columns 2 and 4). Here we find that workers in their traditional occupation earn 12 percent more per hour than non-traditional workers in the same occupation. Outsiders–and particularly those from socially higher castes–appear less productive in the traditional occupations even after controlling for social networks and parental occupations. This result is at odds with the selection of the standard Roy model: with uncorrelated occupational skills, traditional workers should have lower average productivity and hence lower hourly wages than other workers in the same occupation. However, the result is consistent with empirical papers that estimate similar models: Banerjee and Munshi (2004a) find that business-caste "outsiders" perform better in textile manufacturing than traditional farmer castes, while the long-term case study of the Palanpur Village (Lanjouw and Stern, 1998) shows that traditional farmer castes perform better in farming than others. Our structural model extends the standard Roy model to accommodate these empirical patterns.

In Section A2 we discuss a wide variety of alternative empirical specifications, including alternative definitions of networks and fixed effects at the *jati*-state and occupation-state level. Results in Section A5 show that the main results discussed above are robust to all alternatives specifications, with only minor changes in the estimated effect of caste.[23]

Returns to ability in traditional occupations

A possible explanation for these findings could be that traditional occupations–which existed by definition in pre-industrial times–might offer lower returns to ability than non-traditional, "modern" occupations. We test for this by running the following regression:

$$\log(wage/hour)_{iok} = \alpha + \beta TradOcc_{io} + \gamma X_{iok} + \delta TradOcc_{io} * X_{iok} + \varepsilon_{iok},$$

where X_{iok} are years of schooling and experience, which measure workers' general ability. We now define $TradOcc_{io}$ as traditional occupations of *any* jati (hence, it is not indexed by caste k) to

[22] These results provide evidence against the hypothesis that consumers, or intermediaries, have a higher willingness to pay for products sold by traditional workers. If that were the case, workers would likely earn more in their traditional occupation than their fellow caste members in other occupations.

[23] We have further experimented with allowing the effect of paternal occupation to vary with the size of the occupation-caste network. The coefficient on the interacted variable in both occupational choice and wage regressions is always negative, suggesting family and community ties to an occupation may be substitutes. Main results are virtually unchanged by the inclusion of this control. Results available upon request.

examine the characteristics of traditional occupations at the occupation level.[24] All regressions include caste and occupation fixed effects and are estimated separately by gender and conditional on labor force participation.

Column 1 of Table 3 shows that returns to ability are indeed lower in traditional occupations (i.e., $\delta < 0$): each year of schooling increases wages of male workers by 7.2 percent in non-traditional occupations and by only 3.9 percent in traditional occupations. Similarly, returns to experience are more than twice as large in non-traditional occupations. Father's occupation and caste networks increase wages significantly and substantially in magnitude, but these effects do not differ significantly between traditional and non-traditional occupations (Column 2). Including all covariates in Column 3 does not change the results.For women (Columns 4-6), we find much lower and imprecisely measured returns to education and experience in all occupations. Returns to fathers' occupation and caste networks are again large but do not differ significantly between traditional and non-traditional occupations.

These results can rationalize why workers earn higher wages in their traditional occupations compared to other workers in the same occupation (cf., Columns 2 and 4 of 2 which include occupation fixed effects). If workers differ in *general* ability, then high-ability workers sort a priori into modern high-return occupations and low ability workers sort into traditional low-return occupations. The marginal traditional workers who are drawn into their caste's occupation due to the utility boost have therefore higher general ability than the average (non-traditional) worker in the same occupation. Due to their higher general ability, traditional workers can earn more in their traditional occupation than non-traditional workers in the same occupation, as we find in the data. Intuitively, our results indicate that some high ability individuals continue to work in their low-return traditional occupations (e.g., agriculture, laundering, pottery) when, in the absence of the caste-occupation affinity, they might apply their skills more productively in high-return occupations (e.g., teaching, engineering, law).

Discrimination

Next we examine the effects of caste hierarchy on wage discrimination. First, we test for discrimination of "backward" caste groups by including indicators for SCs, STs, and OBCs. Second, we examine discrimination of caste members who work in "superior" occupations by including the variable $OccAboveCaste_{ok}$ which measures the social distance between each respondent's caste and all higher-ranked occupations (i.e., the variable is zero for lower-ranked occupations). In addition, we allow for wage discrimination against women. All specifications control for individual characteristics which include education and experience.

Table 4 shows that wage discrimination is larger for women than for SCs, STs, and OBCs (Column 1). Discrimination of workers in higher-ranked occupations is large and significant (Column 2): a worker whose caste ranking corresponds to the barber/hairdresser profession would earn a 14.4

[24] All results are qualitatively the same, and in some cases more precise, if we restrict our definition of "traditional occupations" to occupations that are traditional for a minimum share of the population (e.g., for more than 0.5 percent).

percent lower wage in clerical work (or similarly ranked occupations). Controlling for discrimination in higher-ranked occupations attenuates, but does not eliminate, the significance of discrimination against backward castes. The estimates are robust to controlling for father's occupation, caste networks, and a traditional occupation dummy (Columns 3 and 4).

5 Model

Guided by these empirical findings, we specify our general equilibrium occupational choice model to study the importance of caste for aggregate outcomes. We first describe the model setup and solve agents' education and occupational choices. We then present the aggregation of individual choices, the production side, and market clearing.

5.1 Model Setup

Individual Characteristics

Individuals i differ in general ability, which consists of an unobservable component α_i and an observable component β_i. Individuals receive an idiosyncratic education cost shock η_i and a vector of idiosyncratic occupation-specific productivity shocks π_{io}.

Caste Affiliation and Family Environment

Individuals i belong to castes k that affect their utility payoffs and choices via four channels. First, workers have a direct preference for working in their caste's traditional occupation and a disinclination for working in occupations that rank socially lower than their caste. We denote these preferences by τ_{ok}. Second, caste members can experience wage discrimination, which we denote by T_{ok}. Third, workers receive productivity effects from their caste network, where we define the network as the share of all workers in an occupation who belong to the worker's caste. Last, caste affiliation can affect costs per year of schooling, which we measure in utils and denote by κ_k. These costs can capture pecuniary costs (e.g., school fees or scholarships) or non-pecuniary factors such as social norms, caste-level discrimination, returns to education in the marriage market, or other factors that can make schooling more or less costly for certain castes.

We allow for productivity effects from working in the same occupation as one's father since parents can transfer skills, customer networks, or other assets to their children.[25] To simplify notation, we denote the total productivity shifter from father's occupation and caste networks for an individual i in occupation o by ψ_{io}.

[25] We assume that intergenerational effects are occupation-specific and only exist if children work in the same occupation as their father. Other intergenerational effects would be subsumed into general ability α_i.

Occupation Characteristics

Each occupation offers a wage rate w_o per human capital unit, which is endogenously determined. Occupations differ in their returns to general ability, which we denote by ρ_o. These returns capture the inherent skill-intensity of an occupation (e.g., engineering is more complex and skill-intensive than agricultural labor). Occupations further differ in amenities A_o which can capture how pleasant it is to work in an occupation as well as entry costs or other attributes that are not directly measurable in wages. In the Indian context, examples of entry costs can include exams to enter government services or high costs of acquiring farm land due to imperfect land markets.

Preferences

Workers have preferences for the homogeneous consumption good C, for the amenities of their occupation A_o, for working in their caste's traditional occupation and against working in socially lower-ranked occupations τ_{ok}. We assume a log-linear functional form, so the utility of a worker i from caste k who works in occupation o is given by:

$$U_{io} = \log(C_{io}) + \tau_{ok} + A_o. \tag{1}$$

The discrete utility τ_{ok} captures the two aspects of the caste system that have been highlighted by anthropologists (cf. Section 2): workers have emotional ties to their traditional occupation and a disinclination for working in occupations that are associated with lower-ranked castes. We allow workers' attachment to traditional occupations to depend on workers' characteristics as we specify below in Section 6.1.

5.2 Education and Occupation Choices

The timing of the model is the following: Individuals live two periods, childhood and adulthood. At birth, they know their caste affiliation, their general ability α_i and their education cost shock η_i. Individuals first choose years of schooling s_i, which remain fixed during adulthood and are a component of workers' observable ability β_i. Young adults then receive idiosyncratic occupation-specific productivity shocks π_{io} and choose an occupation o in which they work during adulthood. Occupation-specific shocks are realized only after education is completed, so children take expectations over these shocks when making their schooling choice. We solve the problem backwards, beginning with the occupational choice.

5.2.1 Occupational Choice

Young adults choose their occupation to maximize utility over their working period of T years, subject to discount factor r, by solving:

$$\max_o \left\{ \int_0^T e^{-rt} \left(\log(C_{io}) + \tau_{ok} + A_o\right) dt \right\}, \tag{2}$$

where C_{io} is consumption, τ_{ok} are workers' preferences for their traditional occupation and against lower-ranked occupations, and A_o are occupational amenities. These amenity shifters allow the model to fit the occupational structure of the economy, including features such as the large mass of workers in seemingly low-returns occupations such as dairy farming. Workers spend their entire income on the final consumption good (which is the numeraire), so that the budget constraint is equal to:

$$C_{io} = (1 - T_{ok}) w_o \Theta_{io}, \tag{3}$$

where T_{ok} is caste wage discrimination, w_o are occupation-specific wage rates, and Θ_{io} are total human capital units that a worker i supplies to occupation o. Workers' human capital units depend on their own characteristics, their father's occupations, and their caste network and are given by:

$$\Theta_{io} = (\alpha_i \beta_i)^{\rho_o} \pi_{io} \psi_{io}, \tag{4}$$

where ψ_{io} captures productivity effects from workers' caste networks and parental occupation, π_{io} is occupation-specific productivity, (α_i, β_i) measure general ability,[26] and ρ_o captures occupation-specific returns to general ability. Modeling unobservable general ability and occupation-specific returns to ability is essential to fit the empirical patterns of occupational choice, selection, and wages that we documented in Section 4 and in Tables 2 and 3. Without these features, our model could not generate the higher wages that we observe for traditional workers relative to outsiders in the same occupation.

Substituting the budget constraint (Equation 3) and the expression for human capital (Equation 4) into the utility maximization (Equation 2) allows us to formulate the occupational choice problem as:

$$\max_o \left\{ \int_0^T e^{-rt} \left[\log \left((1 - T_{ok}) \, w_o \, (\alpha_i \beta_i)^{\rho_o} \, \psi_{io} \right) + \tau_{ok} + A_o + \log(\pi_{io}) \right] dt \right\} \equiv \bar{r} \max_o \left\{ \bar{u}_{io} + \log(\pi_{io}) \right\},$$

where $\bar{r}$ is the discount factor and $\bar{u}_{io}$ is the expected lifetime utility of choosing occupation o (net of the occupation-specific productivity shock). We provide the complete definitions and derivations in Appendix A3.1.

Solving the Occupational Choice Problem

To solve this discrete choice problem, we impose the following assumptions:

Assumption 1. Idiosyncratic productivity shocks $\log(\pi_{io})$ are i.i.d. across occupations and follow a Type-I Extreme Value distribution with location 0 and scale parameter $1/\sigma_\pi$: $\Pr(\epsilon \leq x) = \exp(-\exp(-\sigma_\pi x))$.

[26] Recall that α_i are unobserved ability shocks. The observed component of ability β_i is determined by workers' education and experience. Individuals choose their education during childhood, so it is a fixed characteristic in the occupational choice problem.

Assumption 2. Idiosyncratic ability shocks α_i are i.i.d. across workers and follow a log-normal distribution with mean 0 and variance σ_α^2.

Under Assumption 1, we can express the probability that worker i with ability α_i chooses occupation o as:

$$P_{io|\alpha_i} = \frac{(\exp \bar{u}_{io})^{\sigma_\pi}}{\sum_{o'} (\exp \bar{u}_{io'})^{\sigma_\pi}}. \tag{5}$$

Workers' expected utility before knowing occupation-specific productivity shocks π_{io} is:

$$\mathbb{E}_{\pi_{io}} \left[\bar{r} \max_o \{\bar{u}_{io} + \log(\pi_{io})\} \right] = \frac{\bar{r}}{\sigma_\pi} \log \sum_o (\exp \bar{u}_{io})^{\sigma_\pi}. \tag{6}$$

We then use Assumption 2 to integrate over unobservable ability α_i, so that worker i's unconditional occupational choice probability is equal to:

$$P_{io} = \int P_{io|\alpha_i} \phi(\alpha_i)\, d\alpha_i,$$

where $\phi(\cdot)$ indicates the log-normal PDF. This final integral has no closed-form solution.

5.2.2 Education Choice

Caste affiliation affects education choices through the cost and expected returns to schooling. To capture both of these channels, we augment the standard Mincerian formulation with caste-specific education costs and express individuals' choice of schooling s_i as:

$$\max_{s_i} \left\{ \left(\frac{\bar{r}}{\sigma_\pi} \log \sum_o (\exp \bar{u}_{io})^{\sigma_\pi} \right) - \left(\kappa_{1k} + \frac{\kappa_{2k}}{2} s_i + \eta_i \right) s_i \right\}. \tag{7}$$

The first term of this equation captures returns to schooling through the net present value of expected lifetime utility (derived in Equation 6). When choosing their education, individuals do not yet know their occupation-specific productivity shocks π_{io}; however, they know their occupational choice probabilities and wage payoffs in each occupation, which both depend on caste affiliation and jointly define individuals' expected returns to schooling. An individual who is likely to enter her traditional occupation–in which returns to education are low–invests ex-ante less in education. The second term of Equation 7 represents the cost of education, including caste-specific shifters κ_k and idiosyncratic education cost shocks η_i. In our estimation, we make the following assumption:

Assumption 3. Education cost shocks η_i are i.i.d. across workers and follow a Normal distribution with mean 0 and variance σ_η^2.

Individuals' schooling choice weighs marginal costs against expected marginal returns. We can define the optimal schooling level implicitly by differentiating Equation 7 with respect to s_i. We present the full derivations, including integration and optimization, in Appendix A3.2.

5.2.3 Aggregation of Human Capital in Each Occupation

Individuals' education and occupational choices jointly determine schooling levels, the allocation of general and occupation-specific ability, and occupational caste networks. These factors together determine the total amount of human capital that is supplied to each occupation, which we now derive. First, we use Assumption 1 to solve for workers' expected occupation-specific productivity π_{io} conditional on having chosen occupation o:

$$\mathbb{E}\left(\pi_{io|\alpha_i}\right) = \sigma_\pi \left(\frac{1}{P_{io|\alpha_i}}\right)^{\frac{1}{\sigma_\pi}} \Gamma\left(1 - \frac{1}{\sigma_\pi}\right), \tag{8}$$

where $\Gamma(\cdot)$ is the gamma function. This expression illustrates the negative selection result from the standard Roy model with uncorrelated occupation-specific productivity shocks: within α-types, caste members have an affinity for their traditional occupation which increases their propensity of choosing that occupation ($P_{io|\alpha_i}$) and decreases their average occupation-specific productivity in that occupation. Next, we combine this measure with workers' other characteristics (cf. Equation 4) and we sum across all workers to express total expected human capital in occupation o as:

$$\mathbb{E}\left(\Theta_o\right) = \sum_i \int_{\alpha_i} P_{io|\alpha_i} \left(\alpha_i \beta_i\right)^{\rho_o} \psi_{io} \mathbb{E}\left(\pi_{io|\alpha_i}\right) d\phi\left(\alpha_i\right), \tag{9}$$

where we weigh each observation-type by the corresponding occupational choice probability and use Assumption 2 to integrate over unobservable ability α_i. Across α-types, the negative relationship between castes' occupational choice probability and average human capital is no longer guaranteed. If the utility of working in one's traditional occupation attracts sufficiently high α-types into low-return traditional occupations, then the average human capital of traditional workers in that occupation can be greater than that of outsiders.

5.2.4 Social Networks

We define caste-occupation networks as the share of all workers in an occupation that belong to a given caste, which closely corresponds to earlier work in anthropology (Cohn, 1971)[27] and economics (Munshi, 2011).[28] These networks are a direct function of occupational choices and are given by the following expression:

$$\text{SocialNetwork}_{ok} = \frac{\sum_{i \in k} \int_\alpha P_{io|\alpha_i} d\phi\left(\alpha_i\right)}{\sum_i \int_\alpha P_{io|\alpha_i} d\phi\left(\alpha_i\right)}.$$

[27]Cohn emphasizes that "A particular office or section of a factory may have a high percentage of members of a certain caste or jati, because workers are often recruited for jobs along kin, clan and caste lines. In the Indian context, one is very likely to hire or recommend for hiring someone he knows or can find out about, and one is likely to know more about one's caste fellow than about non caste fellows. (...) The clustering of members of particular jatis can also facilitate cooperation in the work situation, because customs, attitudes and beliefs are similar." Bayly (1999); Deliege (2004); Beteille (2012); Mosse (2020) provide similar arguments.

[28]An alternative measure, the share of each caste that works in an occupation, has the undesirable property that it becomes collinear with measures of traditional occupation as a caste becomes more tied to its traditional occupation. In the extreme case, in which each caste exclusively works in one traditional occupation, this measure of social network would be identical to an indicator for caste's traditional occupation.

Intuitively, a caste that is heavily represented in an occupation can offer its members benefits from increased referrals and greater opportunities for cooperation within that occupation. These productivity effects of social networks imply that workers' occupational choices have important externalities on their fellow caste members. In the absence of a coordinating mechanism, individuals do not internalize these effects and social networks in equilibrium may feature less clustering of castes into the same occupation than in an output-maximizing allocation.

5.3 Firms and Market Clearing

Perfectly competitive firms produce the final consumption good C. The production technology is CES and uses human capital from each occupation Θ_o as inputs. Profit maximization is therefore given by:

$$\max_{\Theta_o} \left\{ A \left[\sum_{o'} Z_{o'} \Theta_{o'}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} - \sum_{o} w_o \Theta_o \right\}, \tag{10}$$

where A is total factor productivity, Z_o is the factor share of each occupation's human capital and σ is the elasticity of substitution between occupations. Firms' first order condition with respect to Θ_o determine human capital demand in each occupation. Wage rates w_o adjust in equilibrium to ensure that labor markets clear, equalizing human capital demand and supply (cf. Equation 9) in each occupation.

5.4 Equilibrium

We formally define the equilibrium in Appendix A3.3. The Appendix further describes how we endogenize wage discrimination T_{ok} by assuming that entrepreneurs experience a disutility from employing workers that belong to certain castes similar to Hsieh et al. (2019).

6 Structural Estimation

We use maximum likelihood to match the model predictions of individuals' wages, education, and occupation choices to their counterparts in our individual-level data. Consistent with the timing assumptions in our model, we estimate two separate likelihood functions: the first for the probability of observed occupational choices and wages, and the second for the probability of observed educational choices. We first describe how we parameterize components of our model and then present the likelihood functions.

6.1 Parameterization and Heterogeneous Effects

Preferences for traditional occupations and against lower-ranked occupations: Individuals' non-pecuniary utility in an occupation is given by::

$$
\begin{aligned}
\tau_{ok} = {} & \mathbb{I}\left(\text{TraditionalOccupation}_k = o\right)\left(\tilde{\tau}_1 + \tilde{\tau}_2 \mathbb{I}\left(\text{OBC}_k\right) + \tilde{\tau}_3 \mathbb{I}\left(\text{SC}_k\right) + \tilde{\tau}_4 \mathbb{I}\left(\text{ST}_k\right) + \tilde{\tau}_5 \mathbb{I}\left(\text{Female}_i\right) + \tilde{\tau}_6 \left(\text{Father's occ.}_i\right)\right) \\
& + \tilde{\tau}_7 \text{LowerOcc}_{ok} + \mathbb{I}\left(\text{Female}_i\right) \times \tilde{\tau}_8 \text{LowerOcc}_{ok} \\
& + \mathbb{I}\left(\text{Homework}_o\right) \times \mathbb{I}\left(\text{Female}_i\right) \times \left(\tilde{\tau}_9 + \tilde{\tau}_{10} \mathbb{I}\left(\text{OBC}_k\right) + \tilde{\tau}_{11} \mathbb{I}\left(\text{SC}_k\right) + \tilde{\tau}_{12} \mathbb{I}\left(\text{ST}_k\right)\right)
\end{aligned}
\tag{11}
$$

Castes' preferences for their traditional occupations are measured by $\mathbb{I}\left(\text{TraditionalOccupation}_k = o\right)$, which indicates whether occupation o is a traditional occupation of caste k. Following the anthropological literature, we allow these preferences to differ by gender, by fathers' occupation,[29] and by castes' social ranking (i.e., SCs, STs, and OBCs). These groups of historically oppressed castes may feel a weaker attachment to their traditional (oftentimes unpleasant and servile) occupations that the caste system imposed upon them.

The anthropological literature further suggests that high castes avoid ritual pollution from working in occupations traditionally associated with socially lower ranked castes. The variable $LowerOcc_{ok}$ measures this disinclination of working in a lower-ranked occupation by quantifying the difference between each caste's social rank and all lower-ranked occupations (i.e., the variable is zero for higher-ranked occupations).

Last, we interact an indicator for home work with a gender dummy since women can face social sanctions when working in the labor market. We allow this effect to vary by caste hierarchy, following evidence that this stigma is strongest for high-caste women (Eswaran et al. (2013) and Cassan and Vandewalle (2021)).

Observable components of general ability: Using the standard Mincer formulation, we parameterize observable human capital β_i as a function of education s_i, experience and experience squared:

$$
\begin{aligned}
\beta_i = \exp & \left(\tilde{\beta}_1 \text{experience}_i + \tilde{\beta}_2 \text{experience}_i^2 + \tilde{\beta}_3 s_i + \right. \\
& \left. + \mathbb{I}\left(\text{Female}_i\right)\left(\tilde{\beta}_4 \text{experience}_i + \tilde{\beta}_5 \text{experience}_i^2 + \tilde{\beta}_6 s_i\right)\right),
\end{aligned}
$$

where we define experience as individuals' age minus their years of schooling s_i minus the typical age to start school (which we set to 6 years).

Productivity effects from social environment: Workers' productivity effects from caste networks and father's occupation in an occupation are given by:

$$
\begin{aligned}
\psi_{io} = \exp & \left(\tilde{\psi}_1 \mathbb{I}\left(\text{Father occ} = o\right) + \tilde{\psi}_2 \text{SocialNetwork}_{ok} + \right. \\
& \left. + \mathbb{I}\left(\text{Female}_i\right)\left(\tilde{\psi}_3 \mathbb{I}\left(\text{Father occ} = o\right) + \tilde{\psi}_4 \text{SocialNetwork}_{ok}\right)\right),
\end{aligned}
$$

[29] Father's occupation could have an additional impact on workers' utility in other occupations. We tested this alternative specification and found this effect to be insignificant for males and slightly negative for women, perhaps due to gender roles. In our preferred specification, we therefore allow father's occupation to only shift workers' utility in their traditional occupation, and their productivity if they work in the same occupation as their father. This specification offers a cleaner identification, since simultaneous effects of parental occupation on their children's utility *and* productivity could only be separately identified through functional form assumptions.

where $\text{SocialNetwork}_{ok}$ is the share of all workers in occupation o that are members of caste k[30] and $\mathbb{I}\,(\text{Father occ} = 0)$ indicates whether the father of individual i worked in occupation o. We allow productivity effects to differ by gender since fathers may differentially transfer their occupation-specific knowledge to sons or daughters, and social networks may be differentially important for women (as shown in Munshi and Rosenzweig, 2006). We set $\text{SocialNetwork}_{ok} = 0$ for home work, since social networks do not seem relevant in this setting.

Wage discrimination: We model wage discrimination in the following way:

$$\begin{aligned}(1 - T_{ok}) \quad = \quad & \exp\left(\tilde{\delta}_1 \mathbb{I}\,(\text{Female}_i) + \tilde{\delta}_2 \text{HigherOcc}_{ok} + \mathbb{I}\,(\text{Female}_i) \times \tilde{\delta}_3 \text{HigherOcc}_{ok}\right. \\ & \left. + \tilde{\delta}_4 \mathbb{I}\,(\text{OBC}_k) + \tilde{\delta}_5 \mathbb{I}\,(\text{SC}_k) + \tilde{\delta}_6 \mathbb{I}\,(\text{ST}_k)\right),\end{aligned} \tag{12}$$

where we allow for discrimination against women and low caste groups (SCs/STs/OBCs) in all occupations. In addition, lower castes were historically barred from certain "superior" occupations (e.g., cooking, or priesthood) and may still face strong discrimination in such high status occupations (as discussed in Section 2). To capture this type of discrimination, we include the $HigherOcc_{ok}$ variable, which quantifies the difference between each caste's social rank and all higher-ranked occupations (i.e., the variable is zero for lower-ranked occupations). We set $T_{ok} = 0$ for home work because caste discrimination is unlikely within the home and because women may face systematic discrimination in all market occupations.

Education cost: Costs per year of schooling κ_k can vary by caste hierarchy (SC/ST/OBC status), gender, and the number of years of schooling:

$$\begin{aligned}\kappa_k \quad = \quad & \tilde{\kappa}_1 + \tilde{\kappa}_2 \mathbb{I}\,(\text{Female}_i) + \tilde{\kappa}_3 \mathbb{I}\,(\text{OBC}_k) + \tilde{\kappa}_4 \mathbb{I}\,(\text{SC}_k) + \tilde{\kappa}_5 \mathbb{I}\,(\text{ST}_k) \\ + \quad & \text{YearsEducation}_i \times \left(\tilde{\kappa}_6 + \tilde{\kappa}_7 \mathbb{I}\,(\text{Female}_i) + \tilde{\kappa}_8 \mathbb{I}\,(\text{OBC}_k) + \tilde{\kappa}_9 \mathbb{I}\,(\text{SC}_k) + \tilde{\kappa}_{10} \mathbb{I}\,(\text{ST}_k)\right).\end{aligned} \tag{13}$$

These costs can capture pecuniary costs as well as group-specific discrimination, social norms, scholarships, reservations, or affirmative action programs. Differences in wealth and financial constraints between low and high castes can further affect education choices, which is also captured in the κ_k parameters.

Occupation parameters: Occupational amenities A_o, wage rates w_o, and skill returns ρ_o are simple vectors where each element represents an occupational category.

Distribution parameters: Last, we estimate the dispersion of the three idiosyncratic shocks: Occupation-specific productivity shocks π_{io} are extreme value distributed with dispersion parameter σ_π; general ability shocks α_i are log-normally distributed with mean 1 and standard deviation σ_α; and education cost shocks η_i are normally distributed with mean 0 and standard deviation σ_η.

[30]We jackknife this variable for the individual's own occupation, subtracting 1 from both the number of caste members and the total workers in the occupation.

6.2 Likelihood Function

With these parameterizations defined, we now turn to our maximum likelihood estimation. The estimation proceeds in two steps: the first for occupation choices and wages, and the second for education choices.

Occupation and wage likelihood

The occupation and wage likelihood function estimates the first set of parameters:
$\Omega_{occ} = \left\{ \tilde{\tau}, \tilde{\beta}, \tilde{\psi}, \tilde{\delta}, A_o, w_o, \rho_o, \sigma_\alpha, \sigma_\pi \right\}$.
The likelihood that a worker i earns wage y_{io} in occupation o can be expressed as the product of the probability that she chooses occupation o and the probability that she earns wage y_{io} conditional on that occupational choice, so that:

$$L_i \left(\hat{y}_{i\hat{o}}, \hat{o}; \Omega, X_i \right) = \int_\alpha \Pr \left[y_{io} = \hat{y}_{i\hat{o}} | o = \hat{o}; \Omega, X_i, \alpha \right] \times \Pr \left[o = \hat{o} | \Omega, X_i, \alpha \right] d\alpha, \tag{14}$$

where X_i are individual characteristics, $\hat{o}$ is chosen occupation, and $\hat{y}_{i\hat{o}}$ is the realized wage in this occupation. We observe these objects in the data. Under the assumption of extreme value distributed productivity shocks, the model admits a closed form expression for occupational choice probabilities (cf. Equation 5):

$$\Pr \left[o = \hat{o} | \Omega, X_i, \alpha_i \right] = P_{io|\alpha_i},$$

and for conditional wage probabilities (derived in Appendix A3.4):

$$\begin{aligned} \Pr \left[y_{io} = \hat{y}_{i\hat{o}} | o = \hat{o}; \Omega, X_i, \alpha_i \right] \quad = \quad & \frac{\sigma_\pi}{\hat{y}_{i\hat{o}}} \left(\frac{\sum_{o'} \left(\exp \bar{u}_{io'|\alpha_i} \right)^{\sigma_\pi}}{\left(\exp \left(\tau_{ok} + A_o + \rho_o \left(\bar{\beta} - \beta_i \right) \right) \hat{y}_{i\hat{o}} \right)^{\sigma_\pi}} \right) \times \\ & \exp \left\{ - \left(\frac{\sum_{o'} \left(\exp \bar{u}_{io'|\alpha_i} \right)^{\sigma_\pi}}{\left(\exp \left(\tau_{ok} + A_o + \rho_o \left(\bar{\beta} - \beta_i \right) \right) \hat{y}_{i\hat{o}} \right)^{\sigma_\pi}} \right) \right\}. \end{aligned}$$

Occupational choice and wage probabilities are both conditional on unobserved ability α_i, over which we integrate in the likelihood function in Equation 14.[31] In the estimation, we normalize the mean and standard deviation of the general α ability distribution for males to one. This normalization does not affect the likelihood value since the mean of unobserved skills is not separately identified from the average wage rates w_o, and the variance of unobserved skills is not separately identified from skill returns ρ_o and coefficients $\tilde{\beta}$.[32] We normalize the occupational amenity A_o in the first occupational category to 1, since occupational utilities are only identified in relative terms.

The wage component of the likelihood is not defined for home workers since they have no observable wage data. We therefore set it to 1 for home workers, using only their occupational choice data to estimate the parameters of home work. It follows that "wages" for home workers are not

[31] We integrate over α using Gauss-Hermite quadrature with 7 nodes.

[32] More generally, an environment with high returns to skill in all occupations and a small variance of skills is observationally equivalent to one with low returns to skill and a large variance of skills.

separately identified from amenities A_o so we normalize $w_{home} = 1$. We do not endogenize the home work "wage" in counterfactuals and we measure output only from market workers.

Education likelihood

Conditional on the first set of parameters, we next use the education likelihood to estimate the remaining parameters: $\Omega_{edu} = \{\tilde{\kappa}, \sigma_\eta\}$. We specify the education likelihood function as a Tobit since almost a third of individuals has no formal education in the data. For these individuals, the schooling choice s_i is likely inframarginal, yielding a likelihood of:

$$L_i(\hat{s}_i) = \int_\alpha \left(\frac{1}{\sigma_\eta}\phi\left(\frac{\hat{\eta}_{i\alpha}}{\sigma_\eta}\right)\right)^{\mathbb{I}(\hat{s}_i>0)} \left(1-\Phi\left(\frac{\hat{\eta}_{i\alpha}}{\sigma_\eta}\right)\right)^{\mathbb{I}(\hat{s}_i=0)} d\alpha, \tag{15}$$

where ϕ is the PDF and Φ the CDF of the standard normal distribution. $\hat{\eta}_{i\alpha}(\hat{s}_i, \hat{y}_{io}, \hat{o}; \Omega)$ are individuals' education cost shocks which rationalize observed education choices $\hat{s}_i$ conditional on individuals' observed wages $\hat{y}_{io}$, occupations $\hat{o}$ and parameters Ω. To characterize these education cost shocks we rearrange the first order conditions of the education choice problem (cf. Equation 7) in the following way:

$$\hat{\eta}_{i\alpha} = -\kappa_{1g} - \kappa_{2g}s_i + \bar{r}\left[-\frac{r}{\sigma_\pi}\log\sum_o \exp(\sigma_\pi \bar{u}_{io}) + \tilde{\beta}_s \sum_o \rho_o P_{io|\alpha_i}\right],$$

where the term in brackets represents expected returns to education during individuals' working period. We provide the full derivation of this expression in Appendix A3.2. As before, we integrate each likelihood contribution in Equation 15 over the distribution of ability shocks α. When solving for the likelihood, we impose the constraint that the second-order conditions of the education choice problem must be negative at the optimal education level to ensure that we derive utility-maximizing education choices.[33]

It is theoretically possible to estimate the occupation-wage and education likelihoods simultaneously. However, it would be computationally infeasible to impose the second order constraint from the education choice problem on the combined likelihood, since the constraint is linear in the education cost κ in the education likelihood, but non-linear in other parameters. We therefore implement the estimation in two steps and bootstrap the standard errors to account for this 2-stage process, clustering at the PSU level.

6.3 Production Parameters

Last, we determine the CES production parameters: occupational intensity Z_o, total factor productivity A, and the elasticity of substitution σ across occupations. We calibrate σ to the literature (setting it to 3) and we compute the other parameters by matching the model's optimality conditions to the data. Dividing firms' first order conditions across two occupations yields the following expression:

[33] Education choices are not well defined in less than 100 out of 688,380 individual×alpha-type combinations as our simulated education choices correspond to local rather than global maxima of the utility function. We interpret this as rejecting the possibility of observing these α values for these particular individuals, and drop them from the estimation.

$$\frac{Z_o}{Z_{o'}} = \frac{w_{o'}}{w_o}\left(\frac{\Theta_o}{\Theta_{o'}}\right)^{\frac{-1}{\sigma}}, \tag{16}$$

where we can compute relative occupational shares Z_o by using our estimated wage rates w_o and by constructing human capital in each occupation Θ_o. The level of occupation shares Z_o is identified because they have to sum to one across all occupations. Using the estimates of σ and Z_o, we then rearrange firms' first order conditions to infer total factor productivity A as:

$$A = \frac{w_o}{Z_o \Theta_o^{\frac{-1}{\sigma}} \left[\sum_o Z_o \Theta_o^{\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\sigma-1}}}. \tag{17}$$

7 Results

7.1 Structural Parameters

We present our maximum likelihood estimates in Tables 5 and 6.

The first two columns of Table 5 present our estimates of workers' preferences for working in their traditional occupation. To provide an interpretation of the estimated parameter values, let us consider their effects on occupational choice probabilities, which are shifted by $\exp(\sigma_\pi \tau_{io})$, as shown in Equation 5. The coefficient on traditional occupation (0.046) implies that general caste men[34] have a 30 (8) percent higher probability of choosing their traditional occupation if their fathers did (did not) work in the occupation. Women are 11 percent less likely to work in the traditional occupation if their fathers did not work in it, and 8 percent more likely if they did. Another perspective comes from comparing the τ_{io} parameters to the variation in occupational amenities A_o, the other non-pecuniary source of occupational utility. Here we see that the τ_{io} shifters are relatively small: the standard deviation of amenities A_o is 3.39 times larger than general caste men's preferences for their traditional occupation. We estimate a significant disutility parameter for working in socially inferior occupations, but only for women. This results can be driven by the fact that many men from elite castes engage in agriculture (a "lower" occupation), but women in those households rarely report work. Relative to general castes, the appeal of the traditional occupation is significantly stronger for STs, and not different for SCs/OBCs. In the first two columns of Table 5 we see that women have a very strong affinity for home work (or, equivalently, stigma for market work). This effect is weaker for castes of lower social status as the literature suggests.

The second pair of columns of Table 5 displays the estimated coefficients $\tilde{\beta}$ that transform years of education and experience into general human capital units β_i. These coefficients are analogous to the standard Mincer coefficients. Averaging the adjusted returns over the occupation-specific return to general human capital ρ_o (shown in Table A9) yields a mean estimate of 0.09, which is consistent with other studies in the Indian context or similar countries (Psacharopoulos and Patrinos, 2004). Experience has positive and nearly linear returns for men, while it has little returns for women.

Column group 4 in Table 5 shows the estimated parameters $\tilde{\psi}$ which determine the productivity

[34] A caste is "general" if it is neither SC, ST nor OBC.

effects from caste-occupation networks and fathers' occupation ψ_{io}. We find strong intergenerational effects: individuals working in the same occupation as their father have a 3.3 times higher productivity than other workers in the same occupation. We also find strong network effects: caste members have a 9 percent higher productivity in an occupation for each 1 percent increase in the share of workers in that occupation who belong to their caste. Consistent with the literature, we find weaker social network effects for women.

The fifth set of columns displays the estimates of wage discrimination $(1 - T_{ok})$. The main victims of discrimination are women, who earn only 25 percent of the wages from identical men in the same occupation. Men working in occupations traditionally associated with castes higher than their own also suffer significant discrimination. For example, a worker from the average barber-status caste would earn 14 percent less due to caste discrimination when working in a white-collar clerical type occupation. We do not find evidence that women suffer from this form of discrimination on top of their baseline discrimination. SCs, STs, and OBCs who work in occupations at their social level, do not seem to experience wage discrimination, instead wages appear even marginally higher for these groups. However, most occupations rank socially higher than these caste groups, so that these individuals receive on net lower wages in a large fraction of potential occupations.

Column group 3 presents the structural coefficients that determine the cost of education. We find that education costs are convex and negative for low years of schooling with costs first decreasing and then increasing after around 6 years of schooling. Costs become positive for schooling levels beyond 9 years for women and beyond 12 years for men. Costs ultimately rise more steeply for SCs and STs. Non-pecuniary rewards (i.e., negative costs) of receiving low education can reflect social stigma, returns on the marriage markets, or compulsory elementary schooling. Schooling choices further depend on idiosyncratic education cost η_i and forgone earnings.

Occupation characteristics: For each occupation, we separately estimate wage rates w_o, amenities A_o, and returns to general human capital ρ_o. We display the full parameter vectors in Appendix Table A9. For each occupation, the wage rate per human capital unit can be interpreted as the intercept of the wage function for individuals with very low human capital. Consistent with this interpretation, the occupations with the highest wage rates ($\ln w_o$) are construction (1.50), and agricultural labor (1.09), and the ones with the lowest values are legal professionals (-5.86) and professors/teachers (-7.655). The highest occupational amenities ($\ln A_o$) are for professors/teachers (3.40) and legal professionals (2.25), while brick/glass makers (0.015) and plantation workers (-0.175) have the lowest. Returns to general human capital (ρ_o) are highest for professors/teachers (2.89) and medical professionals (2.18), and lowest for animal farmers (-1.98) and makers of tobacco products (-1.588). We consider it reassuring that many of these estimates are consistent with reasonable beliefs about the nature of different occupations.

Distributional parameters: Table 6 shows the estimates of distributional parameters. We find that the dispersion of occupational-specific productivity σ_π, of general ability σ_α, and of education cost σ_κ is higher for women than for men. These gender differences could be driven by forces outside our model such as household formation or fertility.

7.2 Counterfactual Results

Our counterfactual analysis explores how the Indian economy would differ if we remove castes' ties to their traditional occupations, castes' hierarchical order, or both.[35] The results for these three sets of counterfactuals are presented in Panel A-C of Table 7. In all counterfactuals, we solve for the new equilibrium and allow wages to adjust endogenously. We assume that individuals have perfect information about all observable variables including wages and caste-occupation networks.[36]

In each set of counterfactuals, we evaluate the direct effects of castes through occupational preferences or discrimination as well as the indirect effects through intergenerational learning and caste networks. When studying direct effects, we keep the distribution of fathers' occupations and caste-occupational networks fixed (Columns 1 and 2) to answer the question, "what would happen if attachment to traditional occupations and/or caste hierarchy ceased to exist in the current generation of workers?" Column 2 allows workers' education to adjust to evaluate the effects of removing caste-specific effects on human capital acquisition.

Removing castes' preferences for traditional occupations (Panel A) has only very minor direct effects (Columns 1 and 2). Market output increases by 0.2 percent with fixed education and by 0.3 with endogenous education. Output per worker increases slightly more and labor force participation drops because some workers (especially women) leave the labor force since they no longer receive a positive utility of working in their traditional occupations. Why is the direct impact of removing traditional occupation affinity so small? First, as mentioned above, the magnitude of the preference parameters $\tilde{\tau}$ is small relative to the variation in other structural parameters such as amenities A_o and productivity effects from networks and fathers' occupation ψ_{io}. The basic structure of the economy therefore remains relatively unchanged: employment shares drop by at most 0.6 percentage points for any occupation, even if the share of traditional workers drops by 5.1 percentage points for the most affected occupation and by 7.9 percent (1 percentage point) in the aggregate. Traditional workers get replaced by other (similar) workers, keeping the occupational structure and aggregate output similar.[37] Despite the trivial aggregate effects, we see improvements in workers' selection

[35]To remove castes' attachment to their traditional occupation, we set $\tilde{\tau}_1 = \tilde{\tau}_2 = ... = \tilde{\tau}_6 = 0$ in Equation 11. To remove caste hierarchy, we first eliminate caste-specific wage discrimination by setting $\tilde{\delta}_2 = \tilde{\delta}_3 = ... = \tilde{\delta}_6 = 0$ in Equation 12. Second, we remove castes' disutility from working in lower-ranked occupations by setting $\tilde{\tau}_7 = \tilde{\tau}_8 = 0$ in Equation 11. Third, we remove differences between castes in women's preference for homework by setting $\tilde{\tau}_{10} = \tilde{\tau}_{11} = \tilde{\tau}_{12} = 0$ and by adjusting $\tilde{\tau}_9$ to be the weighted average of $\tilde{\tau}_9, \tilde{\tau}_{10}, \tilde{\tau}_{11}, \tilde{\tau}_{12}$ in Equation 11. Last, we eliminate caste differences in education costs by setting $\tilde{\kappa}_3 = \tilde{\kappa}_4 = \tilde{\kappa}_5 = \tilde{\kappa}_8 = \tilde{\kappa}_9 = \tilde{\kappa}_{10} = 0$ in Equation 13. The last set of counterfactuals implements all of these changes simultaneously.

[36]Our counterfactuals use the posterior distribution of α_i and the values of η_i generated during the estimation. Thus when simulating occupation choices, years of schooling, and wages at the estimated parameters we generate a baseline very close to the empirical values of these outcomes. We maintain these vectors of unobservables (α_i, η_i) when considering alternative parameters. For individuals with no schooling, we set η_i to the mean of all η_i-values that are consistent with choosing no education. We find nearly identical counterfactual implications when using generic normal distributions for α_i and η_i, however, with greater computational burden.

[37]Banerjee et al. (2013) find similar results for a different aspect of caste identity, individuals' preference to marry within caste: Removing this preference significantly reduces the share of intra-caste marriages, but has only minor effects on marriages and household compositions along other characteristics.

based on occupation-specific productivity π_{io} and general ability (α_i, β_i), as high-ability individuals increasingly select into occupations with high returns to ability ρ_o. However, these gains are partially offset by the fact that fewer workers select into their father's occupation and into the occupation where caste-networks are strongest, which reduces productivity gains from intergenerational learning and networks.

The direct effects of removing caste hierarchy (Panel B) are small with fixed education (Column 1) but larger with endogenous education choices (Column 2) with respective output gains of 0.3 and 3.8 percent. Labor force participation increases (respectively by 0.8 and 0.6 percent) since returns to market employment increases for lower-ranked castes that previously experienced discrimination. Schooling increases by 13 percent because we eliminate caste differences in education costs, and because expected returns to schooling increase due to the removal of wage discrimination in higher-ranked occupations, which oftentimes offer higher skill returns. Allowing education choices to adjust leads to larger changes in the occupational distribution: The most affected occupation sees changes in human capital up to 13.4 percent with endogenous education compared to 3.3 percent when education is fixed. Nonetheless, workers' occupational sorting is still affected by skill transfers in their father's occupation and by caste network effects which we hold constant.

In the next step (Column 3), we additionally eliminate castes' occupational ties through intergenerational learning by removing any correlation between the distribution of father's occupation with either traditional occupations (Panel A), with occupations' hierarchical order (Panel B), or with both (Panel C). To do so, we use a data set at the occupation×individual level and we regress an indicator for father's occupation on an indicator for traditional occupation (or respectively on an indicator of occupations' hierarchical ranking, or on both indicators) and a constant. We then replace fathers' observed occupations with the residual from these regressions–which are by definition orthogonal to traditional occupations, or to castes' and occupations' hierarchical ranking, or to both. To avoid mechanical effects on aggregate output, we rescale the residuals to the same mean as the original father-occupation-indicator.

Removing castes' direct attachment to traditional occupations and this indirect attachment through intergenerational learning (Panel A, Column 3) has large negative effects with a 7 percent decrease in aggregate output. Labor force participation decreases by 5 percent because some workers now choose home production while they were previously attracted to occupations that were simultaneously their father's and traditional occupation. Workers increase schooling and their selection based on individual characteristics. However, output per worker decreases by 2 percent since these productivity gains are smaller than losses from reduced intergenerational learning and lower network effects. Workers sort less towards their traditional occupations, where caste networks (which we still hold constant) are strongest, because they can no longer simultaneously choose their father's and traditional occupation. Distributional effects for castes and occupations are large: the most affected caste sees a 79 percent decrease in average income, the most affected occupation loses 23 percent of its human capital, and the aggregate share of workers who work in their traditional occupations decreases by 32 percent (from 12.1 to 8.3 percent).

Removing direct effects of caste hierarchy and indirect effects through fathers' occupations has smaller but also negative effects with a 2.5 percent decline in aggregate output (Panel B, Column 3). Labor force participation drops by 4.1 percent since workers can no longer choose their father's occupation and the occupations that correspond to their castes' social ranking in which caste networks are strongest. Nonetheless, output per worker increases by 1.7 percent since losses from less intergenerational learning and lower network effects are offset by an expansion in schooling (13 percent) and an improved selection by comparative advantage. Overall, the results in Column 3 of all panels highlight the importance of parental occupation and intergenerational learning for occupational and educational choices and for aggregate outcomes.

The last channel that ties castes to occupations are productivity spillovers from existing networks: once strong caste networks are established in given occupations, they can sustain the selection of caste members into these occupations. We remove these effects by adjusting caste-occupation networks endogenously from occupational choices (Column 4). For this counterfactual, we first solve for the baseline steady state by using our estimated parameter values and solving for caste-occupation networks and occupational human capital that are consistent with individuals' utility-maximizing choices and market clearing. We then compare the counterfactual to this baseline steady state. There is a possibility of multiple equilibria due to the productivity spillovers from caste networks. While we have not exhaustively investigated the set of all possible equilibria, we adopt a numerical approach that aims at identifying the equilibrium that is plausibly "closest" to the existing one.[38] We argue that this equilibrium best captures how caste-occupation networks might evolve if attachment to traditional occupations and/or caste hierarchy were to disappear.

Removing castes' direct and indirect attachment to traditional occupations and endogenizing networks (Panel A, Column 4) leads to even larger negative effects: Output decreases by 9.8 percent, output per worker by 4.5 percent and labor force participation by 5.5 percent. These results highlight that castes' ties to their traditional occupation–either direct ties via preferences or indirect ones via parental occupations–play an important role in organizing castes into strong occupation networks. These coordinating elements are important because individuals do not internalize the large externality of their occupational choices on caste networks. Removing the coordinating forces and allowing networks to adjust endogenously leads to occupational networks that are only weakly clustered at the caste level, which lowers output and productivity. Consistent with this, the share of traditional workers decreases substantially: by 4.3 percentage points in the aggregate, by 10.5 percentage points in the most affected occupation, and by 27.6 percentage points in the most affected caste. Despite the negative aggregate effects, there are also positive aspects: Without strong caste networks in low-return occupations, individuals invest more in education (0.8 percent) and are more likely to pursue occupations aligned with their comparative advantage. Changes in the occupational distribution are therefore large with occupations contracting by up to 32 percent and others expanding by up to 7.6 percent in human capital. These positive effects might be even stronger if, in the absence of caste identity, individuals could strengthen social networks around

[38]To do this, we start from the baseline human capital distribution and baseline caste networks and we perform only minor updates, changing exogenous parameters in several small steps, when solving for the new equilibrium.

other characteristics absent from our model, such as place of origin or religious sect.

Removing instead caste hierarchy (Panel B) and endogenizing networks decreases aggregate output by 1.9 percent (Column 4), which is less than the 2.5 decline with exogenous networks (Column 3). Output per worker increases by 2.6 percent and schooling by 13.4 percent as workers make more efficient educational and occupational choices after we removed castes wage discrimination in higher-ranked occupations. Allowing caste networks to adjust endogenously has a net positive effect on productivity and output (comparing Columns 3 and 4), because castes are still attached to their traditional occupations, which maintains strong networks in traditional occupations. Ties to traditional occupations are most effective at organizing castes into strong networks as they coordinate each caste into one *single* occupation.[39] Caste hierarchy instead disperses caste workers among many similarly-ranked occupations, thus not offering the same benefits in terms of network coordination while preventing workers from making efficient educational and occupational choices. In addition, the endogenization of networks now allows lower-ranked castes to establish networks in higher-ranked and high-return occupations.

In Panel C, we eliminate caste attachment to traditional occupations and caste hierarchy at the same time. The combined direct effects are somewhat larger than the separate cases with output gains of 0.47 and 4.1 percent (Columns 1 and 2). Removing the correlation of fathers' occupations with castes' traditional occupations and with occupational hierarchy leads to output losses of 3.4 percent which lies between the separate cases (Column 3). The removal of all elements that coordinate castes into occupational networks (Column 4) decreases output by 7.1 percent due to weaker caste networks and lower intergenerational learning, which is only partially offset by an expansion in schooling and better sorting of workers based on their individual characteristics.

To document effects on inequality more systematically, we analyze changes along the entire distribution of human capital.[40] In the baseline, women are under-represented in the bottom half and in the top 10 percent of the human capital distribution while they are over-represented between to 50th and 80th percentile. Low castes, in particular low caste males, are over-represented in the bottom 50 percent and under-represented in the top 10 percent.[41]

To track the destiny of individuals across counterfactuals, we use Growth Incidence Curves (Figure 2), in which the x-axis ranks individuals by percentile of their baseline human capital and the y-axis shows the mean human capital growth rate between baseline and counterfactuals for each baseline percentile.[42] Figure 2 shows the Growth Incidence Curves for the same counterfactuals that we present in Table 7 and discuss above.

[39]Consistent with this, the share of traditional workers now decreases only by 9 percent compared to 25 percent when removing castes' attachment to their traditional occupations.

[40]We define individuals' human capital broadly by including occupation-specific productivity, general ability, education, experience, and productivity effects from social networks and parental learning (cf. Equation 4). We focus on human capital instead of income because it is defined regardless of labor force participation.

[41]We show the composition by gender and caste for each percentile of the human capital distribution in Appendix Figure A1.

[42]These Growth Incidence Curves have been proposed by Ravaillon and Chen (2003) and popularized by Milanovic (2016). Ravaillon and Chen (2003) further discusses that the mean growth rate has preferable properties compared to the growth rate of the mean.

In line with the aggregate results, we find that removing castes' direct preferences for traditional occupations has very small distributional effects (Figure 2a). Additionally removing the correlation between fathers' and traditional occupations has large negative effects on human capital in the bottom half of the distribution which are even larger when networks are endogenized with human capital decreasing around 3-4 percent. These workers with low initial human capital benefit most from the coordinating role of castes' traditional occupation and suffer most from weaker networks and reduced intergenerational learning. Effects on individuals in the top half of the human capital distribution are more nuanced: The 50th to 80th percentiles (in which women are over-represented) see large gains up to 10 percent because the removal of caste ties allows them to choose occupations with higher returns to skills and encourages them to invest more in schooling.

Removing the direct effects of caste hierarchy has small effects on human capital when education is fixed but large positive effects throughout the distribution when education is endogenous (Figure 2b). Human capital increases most in the upper half of the distribution where high-caste women are over-represented in the 50th to 80th percentile. This is because we equalize women's disutility of working in the market across all castes, which reduces the previously higher stigma for market work among high-caste women and encourages them to invest in education and to sort into occupations in which they are most productive. Removing additionally the correlation between fathers' occupation and occupations' hierarchical order and endogenizing networks lowers human capital gains for the bottom half, but increases them for the top half of the distribution relative to the counterfactual which fixes fathers' occupation and caste networks. This result suggests again that individuals with initially lower human capital (bottom half) loose more from reduced parental learning and weaker caste networks than they gain from more efficient educational and occupational choices. The opposite seems true for individuals with initially higher human capital (top half).

Combined effects of removing attachment to traditional occupations and caste hierarchy (Figure 2c) are similar to the counterfactual that removes only caste hierarchy–given that effects in this counterfactual are very large. However, combined effects are smaller–and almost null–for the bottom 20 percent which suffers most from reduced intergenerational learning and weaker caste networks now that all elements are removed which coordinated castes into specific occupational networks.

Overall, distributional effects of removing castes' occupational ties are large, especially when removing caste hierarchy. Individuals in the bottom half of the initial human capital distribution (mostly low-caste men) benefit less than individuals between the 50th and 80th percentile (mostly women) and individuals in the top decile (mostly high-caste men). These results confirm the findings of Munshi and Rosenzweig (2006) who show that low-skill men benefit most from caste network effects in traditional occupations.

8 Conclusion

The effect of social identity on occupational choice has often been highlighted as a potential distortion of human capital allocation and source of economic inefficiency. Examining this question in the context of the Indian caste system, we find mixed evidence. Occupational identity and caste hierarchy have major effects on career choices. Certain occupations in India are still composed primarily of individuals "born" into that occupation, and the average person is more than three times as likely to enter their traditional occupation than any other occupation. Individuals born into castes historically designated as polluting are therefore under-represented in occupations with high wages and high returns to human capital.

However, the direct effects of caste identity on occupational choices have only a small impact on the overall efficiency of the macro-economy. Three forces diminish these distortions. First, working in a parent's occupation increases productivity; and parents are more likely to work in the caste's traditional occupation or in an occupation that corresponds to the caste's social status. Hence, we find that intergenerational learning keeps many workers in occupations linked to their castes (either by traditional occupation, or by social ranking), even if workers' preferences for these occupations is removed. Second, the clustering of castes into specific occupations generates positive social network effects that partially compensate for the misallocation of human capital. Even if individuals may be working in the "wrong" occupations, by doing so they increase the productivity of their caste-mates in that occupation. Third, the misallocation of human capital caused by traditional occupations is primarily limited to the reallocation of low-skill individuals between different low-skilled occupations. Since these individuals are numerous, the magnitudes appear high, but the strength of caste identity is not sufficient to draw many high-skilled individuals out of the "modern workforce".

We find larger effects when we eliminate the connections between caste and occupation that run through parental occupations and social networks. Once the ties of occupational affinity and parentally transmitted human capital are broken, we find that individuals' occupational choices are less clustered at the caste-occupation level, weakening networks and lowering network-based productivity spillovers. These effects are particularly harmful to those at the bottom of the human capital distribution. Overall, these factors lead to a reduction in market output, despite improvements in education and human capital allocation. These results highlight the need of taking a broad view when studying the economic importance of frictions. While the direct effects of occupational preferences are small, their indirect effects via social networks and intergenerational learning are much more important.

An important limitation of this study, inherent in the revealed preference approach to occupational identity, is that we can only identify relative and not absolute values of occupation-linked utility. We can therefore not distinguish between a scenario in which individuals receive positive utility from working in their traditional occupation, versus an alternative in which they receive negative utility from all other occupations. With this caveat in mind, we limit our counterfactual analysis to study effects on income and inequality, leaving the evaluation of welfare effects to future work that uses a different methodology.

Our primary focus on the static effects of occupational identity and hierarchy leaves unexamined several dynamic channels through which traditional identities perpetuate themselves over time. If parents believe that their children will benefit from maintaining the traditional occupation, they may choose the occupation themselves in order to "pass on" the tradition. Alternatively, if a history of caste discrimination imposes financial constraints on lower caste households which prevent them from making the necessary investments to enter certain occupations, then the dynamic effects of hierarchical discrimination may be larger than the static distortions. We capture some of these effects through the occupational preference parameters and caste-group specific education costs; however, the short duration of our data set (the IHDS survey) prevents us from characterizing how caste affects inter-generational dynamics within the household.

Our analysis suggests a possible explanation for the remarkable persistence of occupational identities in the 21st century. If the static economic costs are mild, but individuals receive substantial utility from conforming with social norms, then these norms can persist over long periods of time. This may be one reason why castes' occupational identities have endured despite deep changes in the economic structure of the country–which many thought would lead to the weakening of the caste system (Srinivas, 2003). Our analysis suggests, as Ambedkar (1936) anticipated, that the main costs of identity frictions may be dynamic and occur over the course of structural transformation. For individuals who are attached to disappearing occupations, whether Indian handloom weavers at the turn of the 20th century or American manufacturing workers at the beginning of the 21st, there may be substantial costs to occupational change. We leave the study of these important dynamics to future research.

References

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2017). When should you adjust standard errors for clustering? *NBER Working Paper 24003*.

Aggarwal, A., J. Drèze, and A. Gupta (2015). Caste and the power elite in allahabad. *Economic and Political Weekly*, 45–51.

Akerlof, G. (1976). The economics of caste and of the rat race and other woeful tales. *The Quarterly Journal of Economics 90*(4), 599–617.

Akerlof, G. A. (1980). A theory of social custom, of which unemployment may be one consequence. *The Quarterly Journal of Economics 94*(4), 749–775.

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Alvarez-Cuadrado, F., F. Amodio, and M. Poschke (2020, May). Selection and absolute advantage in farming and entrepreneurship: Microeconomic evidence and macroeconomic implications. *CEPR Discussion Paper DP14269*.

Ambedkar, B. R. (1936). *Annihilation of Caste*. Speech prepared for the annual conference of the Jat-Pat-Todak Mandal of Lahore but not delivered.

Amir, S. (2019). Contempt and labour: An exploration through muslim barbers of south asia. *Religions 10*(11), 616.

Banerjee, A., E. Duflo, M. Ghatak, and J. Lafortune (2013). Marry for what? caste and mate selection in modern India. *American Economic Journal: Microeconomics 5*(2), 33–72.

Banerjee, A. and K. Munshi (2004a, 01). How Efficiently is Capital Allocated? Evidence from the Knitted Garment Industry in Tirupur. *The Review of Economic Studies 71*(1), 19–42.

Banerjee, A. and K. Munshi (2004b). How efficiently is capital allocated? evidence from the knitted garment industry in tirupur. *The Review of Economic Studies 71*(1), 19–42.

Bayly, S. (1999). *Caste, Society and Politics in India: From the Eighteenth Century to the Modern Age*. Cambridge University Press.

Beteille, A. (1996). *Caste in Contemporary India*. Oxford University Press.

Beteille, A. (2012). The peculiar tenacity of caste. *Economic and Political Weekly 47*(13), 41–48.

Bidner, C. and M. Eswaran (2015). A gender-based theory of the origin of the caste system of india. *Journal of Development Economics 114*, 142–158.

Borkotoky, K., S. Unisa, and A. K. Gupta (2015). Intergenerational transmission of education in India: evidence from a nationwide survey. *International Journal of Population Research*.

Bryan, G. and M. Morten (2019). The aggregate productivity effects of internal migration: Evidence from indonesia. *Journal of Political Economy 127*(5), 2229–2268.

Cassan, G. (2019). Affirmative action, education and gender: Evidence from india. *Journal of Development Economics 136*, 51–70.

Cassan, G. and L. Vandewalle (2021). Identities and public policies: Unexpected effects of political reservations for women in india. *World Development*.

Chen, R. and Y. Chen (2011, October). The potential of social identity for equilibrium selection. *The American Economic Review 101*(6), 2562–89.

Cohn, B. (1971). *India: The Social Anthropology of a Civilization*. Prentice-Hall.

Cohn, B. (1987). *An Anthropologist Among the Historians*. Oxford University Press.

Conlon, F. (1981). *The Census of India as a Source for Historical Study of Religion and Caste.* New Delhi:Manohar Publications.

Crooke, W. (1896). *The Tribes and Castes of the North Western Provinces and Oudh*. Calcutta: Office of the Superintendent of Government Printing.

Deliege, R. (1993). The myths of origin of the indian untouchables. *Man*, 533–549.

Deliege, R. (2004). *Les Castes en Inde aujourd'hui*. Presses Universitaires de France.

Desai, S. and R. Vanneman (2015). India human development survey-ii (ihds-ii), 2011-12.

Desai, S., R. Vanneman, and National Council Of Applied Economic Research, New Delhi (2008). India human development survey (ihds), 2005.

Deshpande, R. and S. Palshikar (2008). Occupational mobility: How much does caste matter? *Economic and Political Weekly*, 61–70.

Dirks, N. (2001). *Castes of Mind: Colonialism and the Making of Modern India*. Princeton University Press.

Doron, A. (2013). *Life on the Ganga: Boatmen and the ritual economy of Banaras*. Foundation Books.

Dumont, L. (1970). *Homo Hierarchicus: The Caste System and Its Implications*. London: Weidenfeld & Nicolson.

Eckert, F. and M. Peters (2018). Spatial structural change. Technical report.

Eswaran, M., B. Ramaswami, and W. Wadhwa (2013). Status, caste, and the time allocation of women in rural india. *Economic Development and Cultural Change 61*(2), 311–333.

Gandhi, M. K. Varna and caste. `https://www.mkgandhi.org/my_religion/36varna_caste.htm`. Accessed: 2021-7-28.

Ghurye, G. S. (1961). *Caste, Class, and Occupation.* Bombay: Popular Book Depot.

Gupta, D. (2000). *Interrogating caste: Understanding hierarchy and difference in Indian society.* Penguin Books India.

Headley, Z. (2013). Nommer la caste. ordre social et catégorie identitaire en inde contemporaine. *La Vie des idées*.

Heierstad, G. (2017). *Caste, Entrepreneurship and the Illusions of Tradition: Branding the Potters of Kolkata.* Anthem Press.

Heise, S. and T. Porzio (2021). The aggregate and distributional effects of spatial frictions. Technical report.

Hnatkovska, V., A. Lahiri, and S. B. Paul (2013). Breaking the caste barrier. *Journal of Human Resources 48*(2), 435–473.

Hsieh, C.-T., E. Hurst, C. Jones, and P. Klenow (2019). The allocation of talent and U.S. economic growth. *Econometrica 87*(5), 1439–1474.

Ibbetson, D. (1916). *Panjab Castes.* Lahore: Superintendent, Government Printing.

Iversen, V., A. Krishna, and K. Sen (2017). Rags to riches? intergenerational occupational mobility in India. *Economic and Political Weekly 52*(44).

Iversen, V. and P. Raghavendra (2006). What the signboard hides: Food, caste and employability in small south indian eating places. *Contributions to Indian Sociology 40*(3).

Kitts, E. (1885). *A Compendium of the Castes and Tribes Found in India.* Bombay: Education Society Press, Byculla.

Kranton, R. (2016). Identity economics 2016: Where do social distinctions and norms come from? *American Economic Review 106*(5), 405–409.

Kumar, S., A. Heath, and O. Heath (2002). Changing patterns of social mobility: Some trends over time. *Economic and Political Weekly 37*(40), 4091–4096.

Lanjouw, P. and N. Stern (1998). *Economic Development in Palanpur over Five Decades.* Oxford University Press.

Lee, A. (2019). The origins of ethnic activism: Caste politics in colonial India. *The Journal of Race, Ethnicity, and Politics 4*(1), 148–179.

Mayer, A. (1996). *Caste in an Indian Village: Change and Continuity 1954-1992*. Oxford University Press.

Michelutti, L. (2008). 'We are Kshatriyas but we behave like Vaishyas': diet and muscular politics among a community of Yadavs in North India. *South Asia: Journal of South Asian Studies 31*(1), 76–95.

Milanovic, B. (2016). *Global Inequality: A New Approach for the Age of Globalization*. Harvard University Press.

Mosse, D. (2020). The modernity of caste and the market economy. *Modern Asian Studies 54*(4), 1225–1271.

Munshi, K. (2011). Strength in numbers: Networks as a solution to occupational traps. *The Review of Economic Studies 78*(3), 1069–1101.

Munshi, K. (2019). Caste and the indian economy. *Journal of Economic Literature 57*(4), 781–834.

Munshi, K. and M. Rosenzweig (2006). Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy. *The American Economic Review 96*(4), 1225–1252.

Oh, S. (2021). Does identity affect labor supply? Technical report.

Psacharopoulos, G. and H. A. Patrinos (2004). Returns to investment in education: a further update. *Education Economics 12*(2), 111–134.

Ravaillon, M. and S. Chen (2003). Measuring pro-poor growth. *Economic Letters 78*(1), 93–99.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers 3*(2), 135–146.

Schomaker, M. and C. Heumann (2018). Bootstrap inference when using multiple imputation. *Statistics in medicine 37*(14), 2252–2266.

Schwarz, H. (2010). *Constructing the Criminal Tribe in Colonial India: Acting like a Thief*. Wiley Blackwell.

Shankar, B. P. and M. S. Singhe (2014, May). A study on the influencing factors on the occupational changes among beedi rolling women in kulal community. *Social Science Reporter*.

Singh, K. (1996). *Communities, Segments, Synonyms, Surnames and Titles*, Volume 8. Oxford University Press.

Skorikov, V. B. and F. W. Vondracek (2011). Occupational identity. *Handbook of Identity Theory and Research*, 693–714.

Srinivas, M. N. (1962). *Caste in Modern India*. Asia Publishing House.

Srinivas, M. N. (1994). *The dominant caste and other essays*. Oxford University Press, USA.

Srinivas, M. N. (2003). An obituary on caste as a system. *Economic and Political Weekly 38*(5), 455–459.

Vaid, D. (2012). The caste-class association in india: An empirical analysis. *Asian Survey 52*(2), 395–422.

Vaid, D. (2014). Caste in contemporary india: Flexibility and persistence. *Annual Review of Sociology 40*, 391–410.

Venkatesan, S. (2006). Shifting balances in a 'craft community': The mat weavers of Pattamadai, South India. *Contributions to Indian sociology 40*(1), 63–89.

Wiser, W. H. (1936). *The Hindu Jajmani System*. Lucknow Publishing House.

Young, A. (2014). Structural transformation, the mismeasurement of productivity growth, and the cost disease of services. *The American Economic Review 104*(11), 3635–67.

9 Tables.

Table 1: Traditional Occupation and Occupational Choice

	Probability of occupational choice				
	(1)	(2)	(3)	(4)	(5)
	A. Male (N =2,384,389)				
Occ. is caste's trad. occ.	0.067***	0.065***	0.038***	0.036***	0.021***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Occ. below caste		-0.012***	-0.009***	-0.011***	-0.012***
		(0.001)	(0.001)	(0.001)	(0.001)
Occ. above caste		-0.016***	-0.011***	-0.009***	-0.009***
		(0.001)	(0.001)	(0.001)	(0.001)
Occ. is father's occ.			0.311***	0.311***	0.286***
			(0.004)	(0.004)	(0.005)
Caste-occ. network				0.069***	0.077***
				(0.013)	(0.014)
Occ. is caste's trad. occ. × SC					-0.015***
					(0.004)
Occ. is caste's trad. occ. ×					0.117***
occ is father's occ.					(0.009)
	B. Female (N =2,654,085)				
Occ. is caste's trad. occ.	0.026***	0.025***	0.017***	0.015***	0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Occ. below caste		-0.009***	-0.008***	-0.009***	-0.010***
		(0.001)	(0.001)	(0.001)	(0.001)
Occ. above caste		-0.003***	-0.002**	-0.000	0.000
		(0.001)	(0.001)	(0.001)	(0.001)
Occ. is father's occ.			0.107***	0.107***	0.100***
			(0.004)	(0.004)	(0.005)
Caste-occ. network				0.042***	0.044***
				(0.008)	(0.008)
Occ. is caste's trad. occ. * SC					-0.010***
					(0.003)
Occ. is caste's trad. occ. ×					0.039***
occ is father's occ.					(0.012)
Occ. FE	Yes	Yes	Yes	Yes	Yes

Notes: This table reports results of a linear probability model of occupational choice, using data from all 18-60 year old respondents of the 2011 IHDS. We rectangularize the data set to contain all unique combinations of respondents and occupations. The outcome variable is equal to 1 for respondents' chosen occupation and 0 for all other occupations. "Occ. is caste's trad. occ." indicates that an occupations is traditionally performed by the respondent's caste (if any), as defined in Section 3. "Occ. below (above) caste" measures the difference between a caste's social ranking and lower- (higher-) ranked occupations. Caste-occupation networks are equal to the jackknifed ratio between the number of respondents' caste-mates in an occupation divided by the number of all workers in the occupation. The scheduled caste (SC) dummy indicates whether the respondent's reported caste belongs to the state-level list of scheduled castes. Bootstrapped standard errors clustered at the PSU (village) level, with SEs adjusted to account for imputation of missing parental occupation data. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table 2: Traditional Occupation and Wages

	Log wages in chosen occupation			
	Male		Female	
	(1)	(2)	(3)	(4)
Occ. is own caste's	-0.210***	0.122***	-0.153***	0.121***
trad. occ.	(0.030)	(0.029)	(0.039)	(0.037)
Occ. below caste	0.104***	-0.087**	0.729***	-0.148*
	(0.036)	(0.039)	(0.090)	(0.084)
Occ. is father's occ.	-0.043***	0.050***	0.445***	0.357***
	(0.015)	(0.013)	(0.055)	(0.064)
Caste-occ. network	1.853***	0.472**	4.589***	2.783***
	(0.271)	(0.201)	(0.716)	(0.494)
Jati FE	Yes	No	Yes	No
Occ. FE	No	Yes	No	Yes
R^2	0.18	0.25	0.15	0.22
Observations	48,174	48,174	23,695	23,695

Notes: This table reports results of regressing log wages on caste and individual characteristics, using data from all 18-60 year old respondents of the 2011 IHDS. Wage data is windsorized at the 1st and 99th percentile of the non-negative values within occupation category. The variable "Occ. is caste's trad. occ." indicates that an occupations is traditionally performed by the respondent's caste (if any), as defined in Section 3. Caste-occupation networks are equal to the jackknifed ratio between the number of respondents' caste-mates in an occupation divided by the number of all workers in the occupation. "Caste above occ." measures the difference between a caste's social ranking and that of their occupation, if this is positive.

All specifications include controls for state fixed effects, education, age, experience, rural/urban location, OBC/SC/ST status, religion,and a dummy variable for individuals who do not associate with a caste. Cases of missing father's occupation are imputed using all other covariates plus a measure of fathers' occupation distribution from individuals in the same caste.

Bootstrapped standard errors clustered at the PSU (village) level, with SEs adjusted to account for imputation of missing parental occupation data. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table 3: Returns to Human Capital in Traditional Occupations

	Log wages in chosen occupation					
		Male			Female	
	(1)	(2)	(3)	(4)	(5)	(6)
Years education	0.072***		0.072***	0.027*		0.025*
	(0.006)		(0.006)	(0.015)		(0.014)
Years education×	-0.033***		-0.033***	0.004		0.006
Occ. is any trad. occ.	(0.006)		(0.006)	(0.015)		(0.015)
Experience	0.025***		0.026***	0.004		0.004
	(0.002)		(0.002)	(0.004)		(0.004)
Experience×	-0.020***		-0.020***	0.000		0.001
Occ. is any trad. occ	(0.002)		(0.002)	(0.004)		(0.004)
Father's occ.		0.112	0.111*		0.313	0.307
		(0.071)	(0.066)		(0.250)	(0.249)
Father's occ.×		-0.054	-0.048		-0.048	-0.046
Occ. is any trad. occ		(0.073)	(0.069)		(0.248)	(0.247)
Caste-occ. network		1.975***	1.178*		2.753	2.642
		(0.763)	(0.705)		(1.725)	(1.720)
Caste-occ. network×		-1.114	-0.317		0.347	0.418
Occ. is any trad. occ		(0.778)	(0.707)		(1.794)	(1.779)
Jati FE	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.27	0.26	0.27	0.24	0.24	0.24
Observations	48,174	48,174	48,174	23,695	23,695	23,695

Notes: This table reports results of regressing log wages on caste and individual characteristics, using data from all 18-60 year old respondents of the 2011 IHDS. Wage data is taken from the respondent's highest income occupation, with the 1st and 99th percentiles windsorized. The variable "Occ. is any trad. occ" indicates whether an occupation is traditional for *any* caste, as defined in Section 3. Caste-occupation networks are equal to the jackknifed ratio between the number of respondents' caste-mates in an occupation divided by the number of all workers in the occupation. All specifications include controls for state fixed effects, education, experience, rural/urban location, OBC/SC/ST status, religion, and a dummy variable for individuals who do not associate with a caste. Cases of missing father's occupation are imputed using all other covariates plus a measure of fathers' occupation distribution from individuals in the same caste. Bootstrapped standard errors clustered at the PSU (village) level, with SEs adjusted to account for imputation of missing parental occupation data. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table 4: Wage Discrimination

	Log wages in chosen occupation			
	(1)	(2)	(3)	(4)
Female	-0.322***	-0.322***	-0.312***	-0.312***
	(0.013)	(0.012)	(0.012)	(0.012)
Other backwards caste	-0.070***	-0.042**	-0.035**	-0.038**
(OBC)	(0.018)	(0.018)	(0.017)	(0.017)
Scheduled caste	-0.185***	-0.068***	-0.076***	-0.078***
(SC)	(0.018)	(0.021)	(0.021)	(0.021)
Scheduled tribe	-0.231***	-0.139***	-0.140***	-0.147***
(ST)	(0.035)	(0.034)	(0.034)	(0.033)
Occ. above caste		-0.525***	-0.456***	-0.416***
		(0.060)	(0.060)	(0.061)
Father's occ.			0.142***	0.132***
			(0.018)	(0.018)
Caste-occ. network			0.846***	0.630**
			(0.239)	(0.245)
Occ. is caste's trad. occ.				0.113***
				(0.025)
Occupation FE	Yes	Yes	Yes	Yes
R^2	0.218	0.305	0.306	0.307
Observations	71,872	71,872	71,872	71,872

Notes: This table reports results of regressing log wages on caste and individual characteristics, using data from all 18-60 year old respondents of the 2011 IHDS. Wage data is windsorized at the 1st and 99th percentile of the non-negative values within occupation category. "Occ. is caste's trad. occ." indicates that an occupations is traditionally performed by the respondent's caste (if any), as defined in Section 3. "Occ. above caste" measures the difference between respondent's caste's social ranking and all higher-ranked occupations. Caste-occupation networks are equal to the jackknifed ratio between the number of respondents' caste-mates in an occupation divided by the number of all workers in the occupation.

All specifications include controls for state fixed effects, education, experience, rural/urban location, OBC/SC/ST status, religion, and a dummy variable for individuals who do not associate with a caste. Cases of missing father's occupation are imputed using all other covariates plus a measure of fathers' occupation distribution from individuals in the same caste. Bootstrapped standard errors clustered at the PSU (village) level, with SEs adjusted to account for imputation of missing parental occupation data. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table 5: Structural Parameters: Coefficients

Non-pecuniary utility ($\tilde{\tau}$) (1)		General human capital $(\tilde{\beta})$ (2)		Costs of education ($\tilde{\kappa}$) (3)	
Traditional occupation	0.046	Experience	0.021	Constant	-7.069
	(0.020)		(0.002)		(0.191)
Traditional occupation×	-0.112	Experience^2	-0.000	Females	2.677
female	(0.022)		(0.000)		(0.112)
Traditional occupation×	0.109	Education	0.108	Other backwards	0.693
father's occupation	(0.026)		(0.003)	caste	(0.129)
Traditional occupation×	0.015	Experience×	-0.018	Scheduled caste	1.570
OBC	(0.039)	female	(0.002)		(0.131)
Traditional occupation×	-0.021	Experience^2×	0.000	Scheduled tribe	2.386
SC	(0.046)	female	(0.000)		(0.150)
Traditional occupation×	0.090	Education×	-0.059	Constant×	0.717
ST	(0.020)	female	(0.003)	education	(0.022)
Occupation rank	-0.032			Females×	-0.163
lower than caste	(0.038)			education	(0.017)
Occupation rank	-0.413			OBC×	0.014
lower than caste×female	(0.065)			education	(0.015)
Home work×	2.341			SC×	0.003
female	(0.059)			education	(0.011)
Home work×	-0.088			ST×	-0.005
female×OBC	(0.022)			education	(0.011)
Home work×	-0.322				
female×SC	(0.026)				
Home work×	-0.535				
female×ST	(0.048)				

Occupation-specific human capital $(\tilde{\psi})$ (4)		Labor force discrimination $(\tilde{\delta})$ (5)	
Father's occupation	1.199	Female	-1.393
	(0.023)		(0.047)
Caste's share in	6.548	Occupation rank	-0.504
occupation	(0.270)	higher than caste	(0.051)
Father's occupation	1.555	Occupation rank	0.591
×female	(0.046)	higher than caste×female	(0.069)
Caste's share in	-2.354	OBC	0.016
occupation×female	(0.485)		(0.014)
		SC	-0.010
			(0.018)
		ST	-0.020
			(0.028)

Notes: Parameters are estimated by maximum likelihood as described in Section 6. Standard errors in parentheses clustered at the PSU level.

Table 6: Structural Parameters: Variances

	Parameter value
Occupational wage shocks (σ_π)	1.676
	(0.030)
Occupational wage shocks (σ_π)$\times$	0.034
female	(0.024)
General skills (σ_α)$\times$	0.371
female	(0.016)
Cost of education shocks (σ_κ)	3.248
	(0.106)
Cost of education shocks (σ_κ)$\times$	0.178
female	(0.107)

Notes: Parameters are estimated by maximum likelihood as described in Section 6. Standard errors in parentheses clustered at the PSU level.

Table 7: Counterfactual Results: Effects of Removing Castes' Occupational Ties

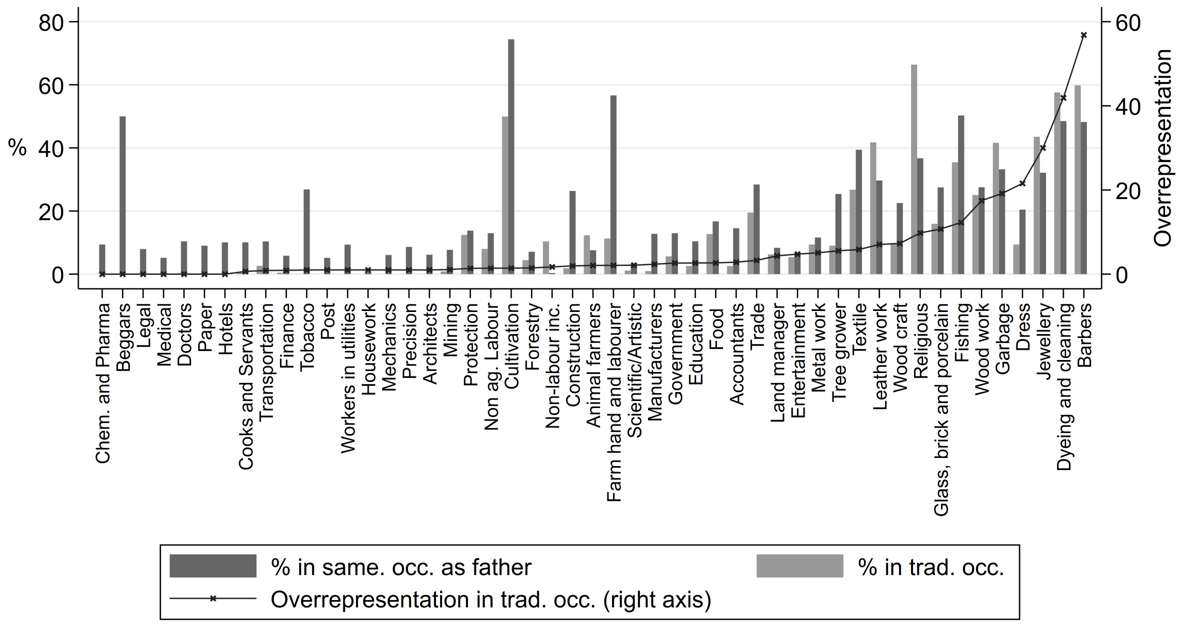
	Eliminating specific caste effects + end. wages	(1) + endogenous education	(2) + parental occupation orthogonal to trad. occ. / to hierarchy of occ. / to both	(3) + endogenous social networks
	(1)	(2)	(3)	(4)
Panel A: Eliminating Caste Ties to Traditional Occupations				
Market Output	0.180	0.309	-7.021	-9.755
Output per worker	0.300	0.427	-2.024	-4.464
Human Capital	0.395	0.756	1.147	-0.208
Labor force participation	-0.119	-0.117	-5.100	-5.538
Schooling		0.093	0.454	0.797
Panel B: Eliminating Caste Hierarchy Effects				
Market Output	0.304	3.838	-2.476	-1.853
Output per worker	-0.520	3.210	1.702	2.567
Human Capital	1.053	12.894	11.968	11.558
Labor force participation	0.828	0.608	-4.107	-4.309
Schooling		12.554	12.843	13.401
Panel C: Eliminating Caste Ties to Traditional Occupations and Hierarchy Effects				
Market Output	0.474	4.143	-3.429	-7.098
Output per worker	-0.235	3.630	1.310	-2.059
Human Capital	1.428	13.721	13.954	12.154
Labor force participation	0.712	0.495	-4.678	-5.145
Schooling		12.642	12.985	13.838

Notes: All numbers represent percent changes from the baseline. Results in Columns 1-3 are relative to a baseline economy that is simulated from the estimated parameter values and the caste network data used in the estimation. Results in Column 4 are relative to a baseline economy with endogenous caste networks for which we solve conditional on all estimated parameters. All counterfactuals use posterior values of α_i and η_i generated during estimation.

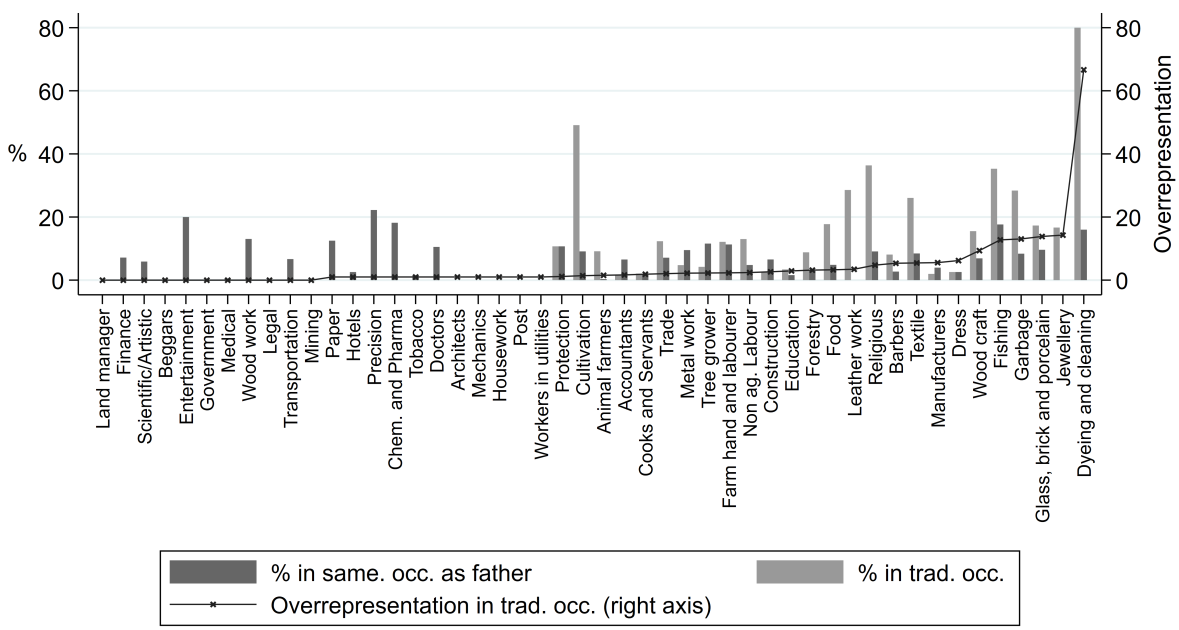
10 Figures

Figure 1: Occupational Composition: Traditional and Parental Transmission

(a) Male Workers



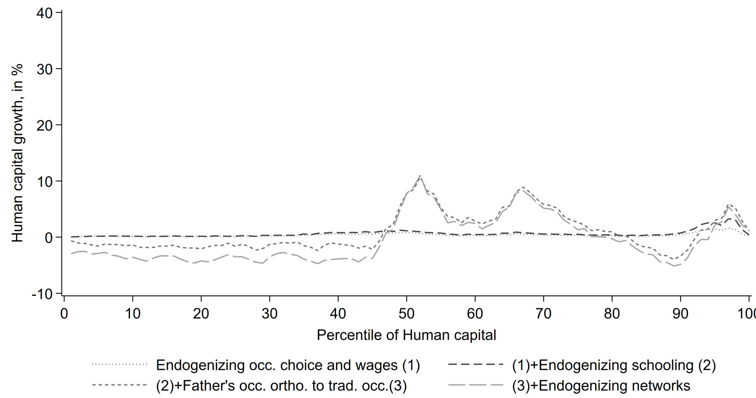
(b) Female Workers



Notes: Panel (a) documents the share of male workers in each occupation who work in their jati's traditional occupation and the share who works in their father's occupation. The line graph shows workers' over-representation in traditional occupations relative to a random allocation. Panel (b) does the same for women. An example of the graph reading is: "60 percent of all male barbers report that profession as their traditional occupation and 50 percent are sons of barbers. Among barbers, there are close to 60 times more barbers from a barber jati than there would be if workers were randomly allocated across the existing occupational structure.

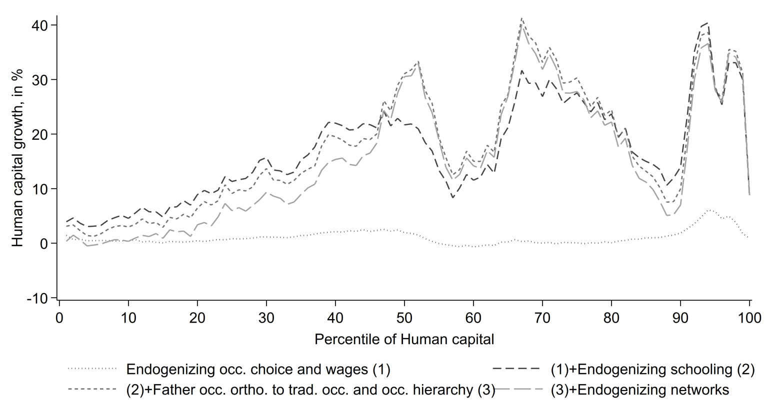
Figure 2: Counterfactuals: Human Capital Growth Incidence Curves

(a) Eliminating Caste Ties to Traditional Occupations (see Table 7, Panel A)



(b) Eliminating Caste Hierarchy Effects (see Table 7, Panel B)

(c) Eliminating Caste Ties to Traditional Occupations and Caste Hierarchy Effects (see Table 7, Panel C)



Notes: This figure shows Growth Incidence Curves in human capital to document the distributional effects of our counterfactuals. The x-axis ranks workers by percentile of their baseline human capital. For each counterfactual, the y-axis shows the mean growth rates of human capital for workers in each percentile. Panel (a) shows the effects of removing caste ties to traditional occupations, Panel (b) of removing caste hierarchy, and Panel (c) of removing both. An example for the interpretation of Figure 2c is: individuals in the bottom 20 percent of the baseline human capital distribution experience on average 0 to 5 percent growth in their human capital depending on the counterfactual.

Appendix

A1 Data Appendix

We restrict our sample to individuals aged 18 to 60 and drop full-time students and unemployed individuals. We drop individuals from the far North-Eastern states of India, Arunachal Pradesh, Nagaland, Manipur, Tripura, Meghalaya, as well as the former Portugese territory of Goa, because the cultural norms in these areas differ significantly from the rest of India.

The IHDS records household and individual income from a wide variety of sources. The survey provides time spent and income earned at the individual level for most occupations; however, we have to make assumptions in some cases.

First, some occupations report income only at the household level (e.g., household businesses), but time spent at the individual-level. We attribute the same hourly wage to each individual, because we do not know individuals' productivity.

Second, for "animal care", the IHDS does not record time spent but instead reports how many animals of each type are owned by the household, and which household members take care of the animals. To derive information on hours worked, we use an additional household survey, the REDS 2006, which is representative for rural India. The survey contains information on time spent on animal care by household members and on the number and types of animals owned by a household. We predict the time that IHDS household members spent in animal care by using the coefficient of an OLS regression of the time spent on animal care of REDS household members on the number of each type of animal and their squared term.

Third, the IHDS records income only at the household level and does not provide time spent for some occupations (e.g., money lending and land rental). We attribute income from these activities to the head of the household if he is younger than 60 years and to his eldest son in the household otherwise. To infer time spent on these activities, we use information on time spent and income earned by the same individual in other occupations, assuming that the share of time spent equals the share of income across occupations (i.e., assuming that individuals have the same average productivity across occupations).

Fourth, we find that almost 34 percent of our sample, mostly women, report their "primary activity" to be housework. However, many of these women indicate that they also spend many hours working for income. We therefore classify a respondent's main occupation to be housework only if she works less than the median number of hours in all other reported occupations (or if she reports no other work). Otherwise we assign her to the occupation in which she earns the highest income. With this procedure, the final sample classifies 12.3 percent of respondents as home workers.

Fifth, many individuals report multiple sources of income. We therefore classify individuals' occupation as the activity in which they earn highest income and spend most time. If these two definitions are different, we choose the occupation that the survey identifies as the "main activity" of the household, and then the activity in which individuals' spend most time.

Sixth, we windsorize hourly wages at the 1st and 99th percentiles. Results are robust, and in fact stronger, when we instead trim at the 0.1st and 99.9th percentiles.

Social and Occupational Rankings

The 1901 Census sorts castes into ranked categories according to social status within province. Provinces differ in the number of categories (six to twelve) and the specific criteria for allocating castes to categories. We quantify these rankings by assuming that the hierarchical categories are uniformly distributed across the range of possible social status (within province) and share the same range across provinces, which we normalize to $(0, 1)$. Thus a caste whose category ranks three out of ten receives a score of 0.3, while another whose category ranks six out of eight scores 0.75.

We then merge these harmonized rankings with the IHDS data for each caste name. The 1901 ranking does not include all castes, so this procedure assigns social ranks to 69 percent of IHDS observations. For the remaining observations, we average the caste rankings of all individuals within the IHDS-generated caste categories in the state (Brahman, OBC, SC, ST, Other) which are known for all individuals. We then assign the remaining 31 percent of unmatched observations the average social rank of the IHDS caste category associated with them in the data.

A2 Additional empirical results

A2.1 Robustness of occupational choice

We perform a variety of robustness checks to confirm the importance of traditional occupations for occupational choices. In Table A1, we show that our main occupational choice analysis is robust to the inclusion of additional controls. Columns 1 and 3 reproduce our main results from Table 1. Columns 2 and 4 additionally control for whether the occupation is agricultural, for inheritance of land, and for intrafamilial transfers of human capital from uncles who live in the same household and work in the same occupation (in addition to the skill transfers from father to child).[43] If traditional occupation is a proxy for inherited physical capital (such as land) or human capital transfers from more distant family members, we might expect these controls to significantly attenuate the traditional occupation coefficient. However, these additional controls do not change our finding that individuals are more likely to work in their jati's traditional occupation.

A2.2 Robustness to network definition

Our preferred measure of network computes the share of a given jati among the workers in an occupation in all of India. In this section, we show that our main reduced form results are unaffected by alternative definitions of networks.

A first concern could be the "reflection problem" inherent in the measurement of networks. In order to alleviate such concerns, we compute two alternative network measures: first, excluding

[43]We identify uncles using the IHDS household roster.

respondent's state of residence; and second, using occupations of the household heads' fathers instead of occupations of the current generation.

A second concern is that our network measure may proxy for other characteristics. Given the village level sampling frame in the IHDS, it is for example possible that the jati network could overlap with the family and/or village network, at least for small jatis. We therefore compute alternative networks using only the NFHS data, while our main regression analysis uses the IHDS sample.

Third, it is possible that local networks are relevant, so we compute networks at the state level instead of the national-level.

Finally, we define networks as the share of a jati among the workers in an occupation instead of the share of a jati working in an occupation. These two definitions capture different views about the role of networks: in the first definition, an individual wants to join an occupation, and considers his network within that occupation. In the latter, an individual looks within her network and makes her occupational choice based on the ease of finding someone in that occupation. We pursue the former as it is more consistent with the literature[44] and because the latter definition (i.e., the share of each caste that works in an occupation) has the undesirable property that it becomes collinear with measures of traditional occupation as a caste becomes more tied to its traditional occupation. In the extreme case, in which each caste exclusively works in one traditional occupation, this measure of social network would be identical to an indicator for caste's traditional occupation.

Tables A2, A3, A4 and A5 present our main regression analysis, using first our preferred network measure and then these alternative network measures. The coefficient on "traditional occupation" remains similar in all specifications except when networks are defined as the share of each jati that works in an occupation. In this case, the coefficient changes in magnitude (and sometimes even switches sign) which we believe to be due to the collinearity between this network measure and the indicator for traditional occupations. The coefficient on network also remains qualitatively comparable across most specifications. In some cases, the coefficient is less precisely estimated, which is likely due to the smaller sample size on which some of the alternative network measures are estimated. In some specifications, the network coefficient decreases in magnitude, which can be due to the lower relevance of the particular network measures (e.g., fathers' network in women's occupational choice regression) or due to collinearity issues (e.g., share in jati network measure in men's wage regression).

A2.3 Robustness to Inclusion of Fixed Effects

Our reduced form results are robust to the inclusion of fixed effects at the state level. We show this in Tables A6 and A7 by including occupation-state fixed effects in the occupational choice regressions and jati-state or occupation-state fixed effects in our wage regressions.

[44] The anthropological literature emphasizes that caste concentration within an occupation is valuable for the hiring of new workers and for productivity in that occupation (cf. Cohn (1971)). Similarly, economic studies of caste and occupation have also measured networks by the share of individuals in an industry who belong to a given caste (Banerjee and Munshi, 2004b; Munshi, 2011).

A2.4 Robustness to imputation of parental occupation

The IHDS data reports the occupation of each household head's father, but not of his mother. Since most household heads are male, we have very little information on father's occupation for women. In addition, mother's occupation may be more relevant for women than father's. In our main analysis, we replace missing data on father's occupation with imputed values based on all variables in the estimating equation, plus the distribution of fathers' occupations in the same caste. Standard errors are adjusted using multiple imputation.

To test if this imputation affects our results, we use the National Election Survey (NES), which was collected by the Center for the Study of Developing Societies (CSDS) in 2009 and 2014.

The NES reports occupations of all respondents (male and female) and of both of their parents, allowing us to measure parental occupation more precisely. However, the NES data has two important drawbacks and differences compared to the IHDS: First, the NES data is collected from electoral rolls so that it is representative of all registered voters but not necessarily of India as a whole. Second, respondents' jatis–and therefore their traditional occupations and their jati-occupation networks–are less precisely measured in the NES. The CSDS cleans and categorizes jati names in the NES; however, this jati categorization is much more aggregated than the one that we use in the main part of our paper. To compute jati-occupation networks, we combine the NES data with IHDS and DHS. To make the jati categorization comparable across the three surveys, we aggregate our jati data from IHDS and DHS to the broader categories provided in the NES.

Table A8 reproduces the occupational choice regressions with NES data, combining the 2009 and 2014 samples and keeping only respondents between ages 18-64 for whom their jati is identified (cf. Table A1 for our main results with IHDS data). The coefficients on "Occ. is father's occ." are qualitatively similar to our main results for both genders, but increase in magnitude. Including mother's occupation does not affect the coefficient on father's occupation for men. For women, the coefficient on father's occupation decreases but remains positive and significant when controlling for mother's occupation (with the size of the coefficient remaining similar to our main results). The coefficient on caste-occupation networks is substantially reduced for both genders–likely because caste categories are much more aggregated compared to our main analysis. The coefficient on "Occ. is caste's trad. occ." is positive and significant but smaller in all specifications–again, this is likely due to the more aggregated jati categorization.

Overall, the NES data with precise measures of parental occupation is able to qualitatively reproduce our main occupational choice results, despite other shortcomings and differences between the NES and IHDS. We therefore conclude that the imputation of parental occupations in the IHDS does not seem to qualitatively alter our main results.

A3 Model Appendix

A3.1 Expected lifetime utility

We follow Mincer's original work and assume that individuals work for T years after finishing school. Expected lifetime utility at the time of the schooling choice is the discounted sum of utility during individuals' working period, starting after their schooling is completed ($t = s$) and ending after T years in the labor market (i.e., $s + T$), so that:

$$U_i^* = \max_s \left\{ \mathbb{E}_{\pi_{io}} \left[\max_o \left\{ \int_s^{s+T} e^{-rt} \left(\log \left((1 - T_{ok}) \, w_o \psi_{io} \left(\alpha_i \beta_i \right)^{\rho_o} \pi_{io} \right) + \tau_{io} + A_o \right) dt \right\} \right] \right\}.$$

We express observable human capital β_i as a function of education s_i and a quadratic function of experience:

$$\beta_i = \exp \left(\tilde{\beta}_s s_i + \tilde{\beta}_x^1 \left(t - s_i - b \right)^1 + \tilde{\beta}_x^2 \left(t - s_i - b \right)^2 \right),$$

where $(t - s_i - b)$ is individuals' experience equal to individuals' age t minus their years of schooling s_i and minus the age at which individuals typically begin school, which we set to 6 and denote by b. The $\tilde{\beta}_s$ and $\tilde{\beta}_x$ coefficients are parameters that map years of schooling and experience into human capital units.

Integrating over years of expected labor force participation yields:

$$\begin{aligned} U_i^* &= \bar{r} \mathbb{E}_{\pi_{io}} \left[\max_o \left\{ \log \left((1 - T_{ok}) \, w_o \psi_{io} \left(\alpha_i \bar{\beta}_i \right)^{\rho_o} \pi_{io} \right) + \tau_{io} + A_o \right\} \right] \\ &\equiv \bar{r} \mathbb{E}_{\pi_{io}} \left[\max_o \left\{ \bar{u}_{io} + \log \left(\pi_{io} \right) \right\} \right], \end{aligned}$$

where $\bar{r} = \frac{e^{-rs}}{r} \left(1 - e^{-rT} \right)$ is the discount factor that depends on years spent in school. We define $\bar{\beta}_i = \exp \left(\tilde{\beta}_s s_i + \bar{\beta}_x \right)$ where $\bar{\beta}_x$ is the pre-employment expected value of experience which is equal to:

$$\bar{\beta}_x = -\tilde{\beta}_x^1 b + \tilde{\beta}_x^2 b^2 + \left(\left(1 - e^{-rT} - e^{-rT} rT \right) \left(\tilde{\beta}_x^1 r + \tilde{\beta}_x^2 \left(-r 2b + 2 \right) \right) - e^{-rT} r^2 T^2 \tilde{\beta}_x^2 \right) / \left(1 - e^{-rT} \right) r^2.$$

A3.2 Educational Choice

Children choose their years of schooling s_i to maximize discounted lifetime utility net of schooling costs. Occupation-specific productivity shocks π_{io} are not known when choosing education, so individuals form expectations about their future occupational choice probabilities based on their other individual characteristics, their caste affiliations, and their parental occupation. Children therefore solve:

$$\begin{aligned} V_i^* &= \max_s \left\{ \bar{r} \mathbb{E}_{\pi_{io}} \left[\max_o \left\{ \bar{u}_{io} + \log \left(\pi_{io} \right) \right\} \right] - \left(\kappa_{1k} + \frac{\kappa_{2k}}{2} s_i + \eta_i \right) s_i \right\} \\ &= \max_s \left\{ \frac{\bar{r}}{\sigma_\pi} \log \sum_o \exp \left(\sigma_\pi \bar{u}_{io} \right) - \left(\kappa_{1k} + \frac{\kappa_{2k}}{2} s_i + \eta_i \right) s_i \right\}, \end{aligned}$$

which yields the following first order condition:

$$(\kappa_{1k} + \kappa_{2k}s_i + \eta_i) + \frac{\bar{r}}{\sigma_\pi} r \log\left[\sum_o (\sigma_\pi \bar{u}_{io})\right] = \bar{r}\tilde{\beta}_s \sum_o \rho_o P_{io}.$$

Individuals choose their level of schooling to equate marginal costs of schooling (left side of the equation) with marginal returns (right side of the equation). Education costs include direct costs (κ and η) and opportunity costs from foregone income. Returns to education depend on the "general" returns to schooling $\tilde{\beta}_s$ multiplied by the probability weighted occupation-specific returns to human capital ρ_o.

A3.3 Equilibrium

To define the equilibrium, we assume that entrepreneurs experience a disutility from hiring certain castes in certain occupations, which we denote by δ_{ok}. This disutility generates wage discrimination T_{ok} which affects castes' effective occupational wage rate per human capital unit. The full set of exogenous parameters is therefore given by: $\Omega = \left\{\delta_{ok}, \tilde{\beta}, \tilde{\psi}, \rho_o, A, Z_o, \sigma, A_o, \tau_{ok}, \kappa_k, \sigma_\pi, \sigma_\alpha, \sigma_\eta\right\}$. Parameters $\{\tilde{\beta}, \tilde{\psi}, \rho_o\}$ determine worker i's productivity in each occupation; $\{A, Z_o, \sigma\}$ characterize the aggregate production function; $\{A_o, \tau_{ok}, \kappa_k\}$ define individuals' utility; and $\{\sigma_\pi, \sigma_\alpha, \sigma_\eta\}$, are dispersion parameters of idiosyncratic productivity shocks. Given these exogenous parameters Ω, the equilibrium of the economy is characterized by:

1. Occupational choice probabilities P_{io} that maximize individuals' utility (cf. Equation 5).

2. Education choices s_i that maximize individuals' utility (cf. Equation 7).

3. Human capital demand Θ_o in each occupation that is consistent with firms' profit maximization (cf. Equation 10).

4. Wage discrimination that exactly offsets entrepreneurs' disutility of hiring certain castes, so that $T_{ok} = \delta_{ok}$. Entrepreneurs are then indifferent between hiring workers from any caste.

5. Wage rates per human capital unit w_o that clear labor markets in each occupation, ensuring that human capital demand equals supply (cf. Equations 10 and 9).

6. Good market clearing, ensuring that total consumption equals total output.

A3.4 Derivation of the wage distribution and likelihood

We now derive the likelihood functions for observed occupations, wages, and schooling levels. We proceed in two steps: first, we derive the distribution of occupation-specific productivity shocks conditional on having chosen an occupation. Second, we build on this result to derive the distribution of workers' income conditional on their occupational choice.

Let V_i^* be the maximum utility of a worker who chooses occupation o^* before his occupational choice:

$$V_i^* = \max_o [V_{io}] = \max_o [\bar{u}_{io} + \log(\pi_{io})] = \bar{u}_{io}^* + \log(\pi_{io}^*).$$

We assume that $\log(\pi_{io})$ is Gumbel distributed. The maximum utility level V_i^* is therefore also Gumbel distributed, so that:

$$\begin{aligned}
\Pr(V_i^* \leq x) &= \Pr(\bar{u}_{io} + \log(\pi_{io}) \leq x)\,\forall o \\
&= \prod_{o'} \exp\{-\exp(-\sigma_\pi(x - \bar{u}_{io'}))\} \\
&= \exp\left\{-\exp\left(-\sigma_\pi\left[x - \frac{1}{\sigma_\pi}\log\sum_{o'}\exp(\sigma_\pi \bar{u}_{io'})\right]\right)\right\},
\end{aligned}$$

where the expression in the last row corresponds to the CDF of the Gumbel distribution with location parameters $\frac{1}{\sigma_\pi}\log\left(\sum_{o'}\exp(\sigma_\pi \bar{u}_{io'})\right)$ and shape parameter σ_π. Using this result, we can now derive the distribution of occupation-specific productivity shocks π_{io}^* for individuals who have chosen occupation o:

$$\begin{aligned}
H_i(x) = \Pr(\pi_{io}^* \leq x | V_{io} = V_i^*) &= \Pr\left(\frac{\exp(V_i^*)}{\exp(\bar{u}_{io}^*)} \leq x\right) \\
&= \exp\left\{-\exp\left(-\sigma_\pi \log\left[x \exp(\bar{u}_{io}^*)\right] + \log\left[\sum_{o'}\exp(\sigma_\pi \bar{u}_{io'}^*)\right]\right)\right\} \\
&= \exp\left\{-x^{-\sigma_\pi}(P_{io}^*)^{-1}\right\}.
\end{aligned}$$

This expression shows that productivity shocks in the chosen occupation π_{io}^* are Frechet distributed with mean $(P_{io}^*)^{-1}$. We now use this result to derive the distribution of workers' income y_{io}^* in their chosen occupation. Recall that workers' income is given by: $y_{io} = (1 - T_k)\, w_o\, (\alpha_i\beta_i)^{\rho_o}\, \psi_{io}\pi_{io}$, so that:

$$\begin{aligned}
J_i(x) &= \Pr(y_{io}^* \leq x | V_{io} = V_i^*) = \Pr(y_{io}^* \leq x) = \Pr\left((1 - T_k)\, w_o\, (\alpha_i\beta_i)^{\rho_o}\, \psi_{io}\pi_{io}^* \leq x\right) \\
&= \exp\left(-\left(\frac{x}{(1 - T_k)\, w_o\, (\alpha_i\beta_i)^{\rho_o}\, \psi_{io}}\right)^{-\sigma_\pi}(P_{io}^*)^{-1}\right) \\
&= \exp\left(\frac{-\sum_o(\exp(\bar{u}_{io'}^*))^{\sigma_\pi}}{\left(\exp\left(\tau_{io} + A_o + \rho_o\left(\bar{\beta}_i - \beta_i\right)\right)x\right)^{\sigma_\pi}}\right).
\end{aligned}$$

Last, we take the derivative of this expression to obtain the PDF of workers' income in their chosen occupation:

$$\begin{aligned}\Pr(y_{io}^* = x | V_{io} = V_i^*) &= \frac{d}{dx} J_i(x) \\ &= \frac{\sigma_\pi}{x} \frac{\sum_{o'} (\exp(\bar{u}_{io'}^*))^{\sigma_\pi}}{\left(\exp\left(\tau_{io} + A_o + \rho_o \left(\bar{\beta}_i - \beta_i\right)\right) x\right)^{\sigma_\pi}} \exp\left(\frac{-\sum_{o'} (\exp(\bar{u}_{io'}^*))^{\sigma_\pi}}{\left(\exp\left(\tau_{io} + A_o + \rho_o \left(\bar{\beta}_i - \beta_i\right)\right) x\right)^{\sigma_\pi}}\right).\end{aligned}$$

A4 Algorithm

We solve for our counterfactuals with a fixed point algorithm. We first modify parameters (or model objects) according to each counterfactual scenario. When caste-occupation networks are exogenous (fixed), we simply iterate on the human capital distribution across occupations until the distribution is consistent with individuals' optimal education and occupational choices at the occupational wage rates that clear labor markets in each occupation. With endogenous caste-occupation networks, we add a second fixed point where we update caste-occupation networks based on individuals' occupational choices in an outer loop. We then iterate on caste-occupation networks and occupations' human capital until these objects are consistent with our equilibrium definition.

A5 Appendix Tables

Table A1: Robustness: Occupational Choice with Additional Controls

	Probability of occupational choice			
	Male		Female	
	(1)	(2)	(3)	(4)
Occ. is caste's trad. occ.	0.021***	0.021***	0.012***	0.011***
	(0.001)	(0.001)	(0.002)	(0.002)
Occ. is father's occ.	0.285***	0.283***	0.100***	0.100***
	(0.005)	(0.005)	(0.005)	(0.005)
Caste-occ. network	0.077***	0.077***	0.044***	0.043***
	(0.014)	(0.014)	(0.008)	(0.008)
Occ. is caste's trad. occ.×SC	-0.015***	-0.014***	-0.010***	-0.009***
	(0.004)	(0.004)	(0.003)	(0.003)
Occ. below caste	-0.012***	-0.012***	-0.010***	-0.010***
	(0.001)	(0.001)	(0.001)	(0.001)
Occ. above caste	-0.009***	-0.007***	-0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Occ. is caste's trad. occ. ×	0.118***	0.111***	0.038***	0.036***
occ is father's occ.	(0.009)	(0.009)	(0.011)	(0.011)
Agricultural occ.×		0.006***		0.004***
land inherited		(0.001)		(0.001)
Occ. is uncle's occ.		0.120***		-0.013
		(0.023)		(0.030)
Occ. FE	Yes	Yes	Yes	Yes
R^2	0.14	0.15	0.33	0.33
Observations	2,384,389	2,384,389	2,654,085	2,654,085

Notes: This table reports results of a linear probability model of occupational choice, using data from all 18-60 year old respondents of the 2011 IHDS. We rectangularize the data set to contain all unique combinations of respondents and occupations. The outcome variable is equal to 1 for respondents' chosen occupation and 0 for all other occupations. "Occ. is caste's trad. occ." indicates that an occupations is traditionally performed by the respondent's caste (if any), as defined in Section 3. "Occ. below (above) caste" measures the difference between a caste's social ranking and lower- (higher-) ranked occupations. Caste-occupation networks are equal to the jackknifed ratio between the number of respondents' caste-mates in an occupation divided by the number of all workers in the occupation. The scheduled caste (SC) dummy indicates whether the respondent's reported caste belongs to the state-level list of scheduled castes. Cases of missing father's occupation are imputed using all other covariates plus a measure of fathers' occupation distribution from individuals in the same caste.
Bootstrapped standard errors clustered at the PSU (village) level, with SEs adjusted to account for imputation of missing parental occupation data. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table A2: Traditional Occupation and Occupational Choice, Alternative Network Definitions, Males

	Probability of occupational choice					
	(1)	(2)	(3)	(4)	(5)	(6)
Occ. is caste's trad. occ.	0.021***	0.021***	0.021***	0.021***	0.021***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Occ. below caste	-0.012***	-0.011***	-0.012***	-0.012***	-0.011***	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Occ. above caste	-0.009***	-0.009***	-0.008***	-0.009***	-0.010***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Occ. is father's occ.	0.286***	0.286***	0.286***	0.286***	0.286***	0.270***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
Occ. is caste's trad. occ. × SC	-0.015***	-0.016***	-0.015***	-0.015***	-0.015***	-0.011***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
Occ. is caste's trad. occ. ×	0.117***	0.116***	0.118***	0.117***	0.118***	0.088***
occ is father's occ.	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)
Caste-occ. network	0.077***					
	(0.014)					
Caste-occ. network (State)		0.037***				
		(0.002)				
Caste-occ. network (State excl.)			0.101***			
			(0.006)			
Caste-occ. network (NFHS only)				0.070***		
				(0.005)		
Caste-occ. network (Fathers)					0.043***	
					(0.004)	
Caste-occ. network (Share in jati)						0.601***
						(0.012)
Occ. FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,384,389	2,384,389	2,384,389	2,384,389	2,384,389	2,384,389

Notes: This table reports results of a linear probability model of occupational choice, using data from all 18-60 year old respondents of the 2011 IHDS. We rectangularize the data set to contain all unique combinations of respondents and occupations. The outcome variable is equal to 1 for respondents' chosen occupation and 0 for all other occupations. "Occ. is caste's trad. occ." indicates that an occupations is traditionally performed by the respondent's caste (if any), as defined in Section 3. "Occ. below (above) caste" measures the difference between a caste's social ranking and lower- (higher-) ranked occupations. Caste-occupation networks are equal to the jackknifed ratio between the number of respondents' caste-mates in an occupation divided by the number of all workers in the occupation. The scheduled caste (SC) dummy indicates whether the respondent's reported caste belongs to the state-level list of scheduled castes. Cases of missing father's occupation are imputed using all other covariates plus a measure of fathers' occupation distribution from individuals in the same caste.
Bootstrapped standard errors clustered at the PSU (village) level, with SEs adjusted to account for imputation of missing parental occupation data. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table A3: Traditional Occupation and Occupational Choice, Alternative Network Definitions, Females

	Probability of occupational choice					
	(1)	(2)	(3)	(4)	(5)	(6)
Occ. is caste's trad. occ.	0.012***	0.012***	0.012***	0.012***	0.014***	-0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Occ. below caste	-0.010***	-0.009***	-0.009***	-0.009***	-0.008***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Occ. above caste	0.000	-0.000	-0.000	0.000	-0.002**	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Occ. is father's occ.	0.100***	0.100***	0.100***	0.100***	0.100***	0.085***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
Occ. is caste's trad. occ. × SC	-0.010***	-0.010***	-0.009***	-0.010***	-0.009***	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Occ. is caste's trad. occ. ×	0.039***	0.039***	0.039***	0.039***	0.039***	0.007
occ is father's occ.	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)
Caste-occ. network	0.044***					
	(0.008)					
Caste-occ. network (State)		0.020***				
		(0.001)				
Caste-occ. network (State excl.)			0.040***			
			(0.004)			
Caste-occ. network (NFHS only)				0.052***		
				(0.005)		
Caste-occ. network (Fathers)					-0.005*	
					(0.003)	
Caste-occ. network (Share in jati)						0.763***
						(0.023)
Occ. FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,654,085	2,654,085	2,654,085	2,654,085	2,654,085	2,654,085

Notes: This table reports results of a linear probability model of occupational choice, using data from all 18-60 year old respondents of the 2011 IHDS. We rectangularize the data set to contain all unique combinations of respondents and occupations. The outcome variable is equal to 1 for respondents' chosen occupation and 0 for all other occupations. "Occ. is caste's trad. occ." indicates that an occupations is traditionally performed by the respondent's caste (if any), as defined in Section 3. "Occ. below (above) caste" measures the difference between a caste's social ranking and lower- (higher-) ranked occupations. The scheduled caste (SC) dummy indicates whether the respondent's reported caste belongs to the state-level list of scheduled castes. The different measures of caste-occupation networks is described in Section A2.2. Cases of missing father's occupation are imputed using all other covariates plus a measure of fathers' occupation distribution from individuals in the same caste.
Bootstrapped standard errors clustered at the PSU (village) level, with SEs adjusted to account for imputation of missing parental occupation data. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table A4: Traditional Occupation and Wages, Alternative Network Definitions, Males

	Log wages in chosen occupation											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Occ. is own caste's trad. occ.	-0.210***	0.122***	-0.185***	0.105***	-0.205***	0.125***	-0.195***	0.129***	-0.234***	0.132***	-0.124***	0.112***
	(0.030)	(0.029)	(0.029)	(0.028)	(0.030)	(0.029)	(0.034)	(0.032)	(0.034)	(0.032)	(0.030)	(0.030)
Occ. is father's occ.	-0.043***	0.050***	-0.046***	0.048***	-0.040***	0.050***	-0.057***	0.046***	-0.072***	0.055***	-0.023	0.046***
	(0.015)	(0.013)	(0.015)	(0.013)	(0.015)	(0.013)	(0.016)	(0.015)	(0.016)	(0.014)	(0.015)	(0.013)
Occ. below caste	0.104***	-0.087**	0.071**	-0.076*	0.104***	-0.086**	0.020	-0.117**	0.012	-0.096**	0.055	-0.060
	(0.036)	(0.039)	(0.036)	(0.040)	(0.036)	(0.040)	(0.038)	(0.046)	(0.046)	(0.048)	(0.036)	(0.040)
Caste-occ. network	1.853***	0.472**										
	(0.271)	(0.201)										
Caste-occ. network (State)			0.286***	0.377***								
			(0.095)	(0.099)								
Caste-occ. network (State excl.)					2.066***	0.403**						
					(0.303)	(0.185)						
Caste-occ. network (NFHS only)							1.552***	0.308				
							(0.303)	(0.213)				
Caste-occ. network (Fathers)									1.253***	0.074		
									(0.163)	(0.149)		
Caste-occ. network (Share in jati)											-0.453***	0.264**
											(0.126)	(0.112)
Jati FE	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Occ. FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.18	0.25	0.18	0.25	0.18	0.25	0.19	0.25	0.18	0.25	0.18	0.25
Observations	48,173	48,173	48,173	48,173	48,173	48,173	48,173	48,173	48,173	48,173	48,173	48,173

Notes: This table reports results of regressing log wages on caste and individual characteristics, using data from all 18-60 year old respondents of the 2011 IHDS. Wage data is taken from the respondent's highest income occupation, trimming the 1st and 99th percentiles. "Occ. is caste's trad. occ." indicates that an occupations is traditionally performed by the respondent's caste (if any), as defined in Section 3. Caste-occupation networks are equal to the jackknifed ratio between the number of respondents' caste-mates in an occupation divided by the number of all workers in the occupation. "Occ. below caste" measures the (absolute) difference between a caste's social ranking and lower-ranked occupations. The scheduled caste (SC) dummy indicates whether the respondent's reported caste belongs to the state-level list of scheduled castes.
All specifications include controls for state fixed effects, education, experience, rural/urban location, OBC/SC/ST status, religion, missing paternal occupation, and a dummy variable for individuals who do not associate with a caste. The different measures of caste-occupation networks is described in Section A2.2. Cases of missing father's occupation are imputed using all other covariates plus a measure of fathers' occupation distribution from individuals in the same caste.
Bootstrapped standard errors clustered at the PSU (village) level, with SEs adjusted to account for imputation of missing parental occupation data. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table A5: Traditional Occupation and Wages, Alternative Network Definitions, Females

	Log wages in chosen occupation											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Occ. is own caste's trad. occ.	-0.153***	0.121***	-0.132***	0.117***	-0.129***	0.145***	-0.136***	0.113***	-0.199***	0.151***	-0.112***	0.111***
	(0.039)	(0.037)	(0.039)	(0.038)	(0.038)	(0.036)	(0.041)	(0.040)	(0.043)	(0.035)	(0.039)	(0.039)
Occ. is father's occ.	0.445***	0.357***	0.445***	0.359***	0.451***	0.361***	0.438***	0.380***	0.329***	0.358***	0.439***	0.335***
	(0.055)	(0.064)	(0.054)	(0.064)	(0.055)	(0.064)	(0.056)	(0.069)	(0.058)	(0.061)	(0.058)	(0.065)
Occ. below caste	0.729***	-0.148*	0.616***	-0.094	0.700***	-0.132	0.685***	-0.180*	0.567***	-0.059	0.574***	0.009
	(0.090)	(0.084)	(0.084)	(0.084)	(0.090)	(0.086)	(0.092)	(0.095)	(0.102)	(0.093)	(0.083)	(0.084)
Caste-occ. network	4.589***	2.783***										
	(0.716)	(0.494)										
Caste-occ. network (State)			0.603***	0.636***								
			(0.177)	(0.155)								
Caste-occ. network (State excl.)					4.414***	2.096***						
					(0.829)	(0.447)						
Caste-occ. network (NFHS only)							3.742***	2.798***				
							(0.643)	(0.501)				
Caste-occ. network (Fathers)									3.284***	0.895***		
									(0.528)	(0.342)		
Caste-occ. network (Share in jati)											0.188	0.873***
											(0.198)	(0.152)
Jati FE	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Occ. FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.14	0.22	0.14	0.22	0.14	0.22	0.14	0.22	0.15	0.22	0.14	0.22
Observations	23,567	23,567	23,567	23,567	23,567	23,567	23,567	23,567	23,567	23,567	23,567	23,567

Notes: This table reports results of regressing log wages on caste and individual characteristics, using data from all 18-60 year old respondents of the 2011 IHDS. Wage data is taken from the respondent's highest income occupation, trimming the 1st and 99th percentiles. The variable "Occ. is caste's trad. occ." indicates that an occupations is traditionally performed by the respondent's caste (if any), as defined in Section 3. Caste-occupation networks are equal to the jackknifed ratio between the number of respondents' caste-mates in an occupation divided by the number of all workers in the occupation. "Occ. below caste" measures the (absolute) difference between a caste's social ranking and lower-ranked occupations. The scheduled caste (SC) dummy indicates whether the respondent's reported caste belongs to the state-level list of scheduled castes.
All specifications include controls for state fixed effects, education, experience, rural/urban location, OBC/SC/ST status, religion, missing paternal occupation, and a dummy variable for individuals who do not associate with a caste. The different measures of caste-occupation networks is described in Section A2.2. Cases of missing father's occupation are imputed using all other covariates plus a measure of fathers' occupation distribution from individuals in the same caste.
Bootstrapped standard errors clustered at the PSU (village) level, with SEs adjusted to account for imputation of missing parental occupation data. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table A6: Robustness: Occupational Choice with Additional Fixed Effects

	Probability of occupational choice	
	Male	Female
Occ. is caste's trad. occ.	0.021***	0.010***
	(0.001)	(0.002)
Occ. is father's occ.	0.276***	0.093***
	(0.005)	(0.004)
Caste-occ. network	0.076***	0.034***
	(0.014)	(0.006)
Occ. is caste's trad. occ.×SC	-0.012***	-0.002
	(0.004)	(0.003)
Occ. below caste	-0.014***	-0.010***
	(0.001)	(0.001)
Occ. above caste	-0.007***	0.001**
	(0.001)	(0.001)
Occ. is caste's trad. occ. ×	0.117***	0.029***
occ is father's occ.	(0.008)	(0.010)
Individual FE	Yes	Yes
Occ-State. FE	Yes	Yes
R^2	0.15	0.36
Observations	2,391,494	2,661,827

Notes: This table reports results of a linear probability model of occupational choice, using data from all 18-60 year old respondents of the 2011 IHDS. We rectangularize the data set to contain all unique combinations of respondents and occupations. The outcome variable is equal to 1 for respondents' chosen occupation and 0 for all other occupations. "Occ. is caste's trad. occ." indicates that an occupations is traditionally performed by the respondent's caste (if any), as defined in Section 3. "Occ. below (above) caste" measures the difference between a caste's social ranking and lower- (higher-) ranked occupations. Caste-occupation networks are equal to the jackknifed ratio between the number of respondents' caste-mates in an occupation divided by the number of all workers in the occupation. The scheduled caste (SC) dummy indicates whether the respondent's reported caste belongs to the state-level list of scheduled castes. Cases of missing father's occupation are imputed using all other covariates plus a measure of fathers' occupation distribution from individuals in the same caste.
Bootstrapped standard errors clustered at the PSU (village) level, with SEs adjusted to account for imputation of missing parental occupation data. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table A7: Robustness: Traditional Occupation and Wages with Additional Fixed Effects

	Log wages in chosen occupation			
	Male		Female	
	(1)	(2)	(3)	(4)
Occ. is own caste's	-0.204***	0.077***	-0.152***	0.127***
trad. occ.	(0.033)	(0.026)	(0.041)	(0.037)
Occ. is father's occ.	-0.052***	0.034***	0.435***	0.324***
	(0.015)	(0.012)	(0.055)	(0.063)
Caste-occ. network	1.819***	0.665***	5.187***	2.743***
	(0.289)	(0.194)	(0.759)	(0.509)
Occ. below caste	0.114***	-0.088**	0.821***	-0.230**
	(0.039)	(0.040)	(0.097)	(0.093)
Jati-State FE	Yes	No	Yes	No
Occ.-State FE	No	Yes	No	Yes
R^2	0.20	0.29	0.17	0.25
Observations	48,174	48,174	23,695	23,695

Notes: This table reports results of regressing log wages on caste and individual characteristics, using data from all 18-60 year old respondents of the 2011 IHDS. Wage data is taken from the respondent's highest income occupation, trimming the 1st and 99th percentiles. "Occ. is caste's trad. occ." indicates that an occupations is traditionally performed by the respondent's caste (if any), as defined in Section 3. "Occ. below caste" measures the (absolute) difference between a caste's social ranking and lower-ranked occupations. Caste-occupation networks are equal to the jackknifed ratio between the number of respondents' caste-mates in an occupation divided by the number of all workers in the occupation. The scheduled caste (SC) dummy indicates whether the respondent's reported caste belongs to the state-level list of scheduled castes. All specifications include controls for education, experience, rural/urban location, OBC/SC/ST status, religion, missing paternal occupation, a dummy variable for individuals who do not associate with a caste, and state fixed effects. Other fixed effects are included as specified in each column. Cases of missing father's occupation are imputed using all other covariates plus a measure of fathers' occupation distribution from individuals in the same caste.
Bootstrapped standard errors clustered at the PSU (village) level, with SEs adjusted to account for imputation of missing parental occupation data. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table A8: Traditional Occupation and Occupational Choice: NES data

	Probability of occupational choice					
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Male (N =446,918)			
Occ. is caste's trad. occ.	0.055***	0.052***	0.019***	0.019***	0.005**	0.005**
	(0.004)	(0.005)	(0.003)	(0.003)	(0.002)	(0.002)
Occ. below caste		-0.018***	-0.009***	-0.009***	-0.011***	-0.010***
		(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Occ. above caste		-0.012***	-0.006***	-0.006***	-0.007***	-0.007***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Occ. is father's occ.			0.541***	0.540***	0.517***	0.492***
			(0.009)	(0.009)	(0.010)	(0.010)
Caste-occ. network				0.037***	0.042***	0.042***
				(0.012)	(0.014)	(0.013)
Occ. is caste's trad. occ. * SC					0.002	0.003
					(0.006)	(0.006)
Occ. is caste's trad. occ. × occ is father's occ.					0.102***	0.097***
					(0.013)	(0.013)
Occ. is mother's occ.						0.118***
						(0.007)
			B. Female (N =411,160)			
Occ. is caste's trad. occ.	0.023***	0.021***	0.007***	0.007***	0.004**	0.004**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Occ. below caste		-0.015***	-0.010***	-0.010***	-0.011***	-0.006***
		(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Occ. above caste		-0.004***	0.000	0.000	-0.000	-0.002
		(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Occ. is father's occ.			0.250***	0.250***	0.242***	0.120***
			(0.010)	(0.010)	(0.011)	(0.010)
Caste-occ. network				0.019*	0.020*	0.015
				(0.011)	(0.011)	(0.011)
Occ. is caste's trad. occ. * SC					-0.002	-0.004
					(0.006)	(0.005)
Occ. is caste's trad. occ. × occ is father's occ.					0.033*	0.014
					(0.019)	(0.013)
Occ. is mother's occ.						0.529***
						(0.015)
Occ. FE	Yes	Yes	Yes	Yes	Yes	Yes

This Table reports results of a linear probability model of occupational choice, using data from all 18-60 year old respondents of the 2009 and 2014 NES. We rectangularize the data set to contain all unique combinations of respondents and occupations. The outcome variable is equal to 1 for respondents' chosen occupation and 0 for all other occupations. "Occ. is caste's trad. occ." indicates that an occupations is traditionally performed by the respondent's caste (if any), as defined in Section 3. "Occ. below (above) caste" measures the difference between a caste's social ranking and lower- (higher-) ranked occupations. Caste-occupation networks are equal to the jackknifed ratio between the number of respondents' caste-mates in an occupation divided by the number of all workers in the occupation. The scheduled caste (SC) dummy indicates whether the respondent's caste belongs to the state-level list of scheduled castes. C

Table A9: Occupation-level Structural Parameters

	All-India occupation social rank	Occupation skill-wage ($\ln w_o$)	Occupation Amenity ($\ln A_o$)	Returns to skill ρ_o
	(1)	(2)	(3)	(4)
Non-labor income earner	0.710	-2.313 (0.087)	2.101 (0.081)	1.146 (0.032)
Cultivation	0.628	-1.365 (0.043)	2.048 (0.064)	1.172 (0.021)
Land manager	0.732	-3.648 (0.976)	1.664 (0.593)	1.311 (0.323)
Agricultural Laborers	0.286	1.094 (0.041)	0.145 (0.079)	-0.174 (0.075)
Animal farmers	0.499	-1.874 (0.056)	1.805 (0.073)	-1.977 (0.031)
Plantation; Tree and Shrub Crop Growers	0.244	-0.464 (0.152)	-0.175 (0.108)	-0.830 (0.173)
Fish related workers	0.244	-1.050 (0.562)	0.440 (0.372)	0.432 (0.505)
Forest hunters, gatherers and officers	0.380	-0.874 (0.435)	0.387 (0.303)	0.654 (0.209)
Mining related worker	0.332	-1.112 (0.318)	0.272 (0.205)	0.674 (0.182)
Laborers, non-agricultural	0.188	0.681 (0.064)	0.074 (0.077)	0.225 (0.056)
Chemical and pharma related worker	0.208	-2.011 (0.730)	0.559 (0.356)	0.966 (0.276)
Textile related worker	0.308	-0.492 (0.275)	0.532 (0.173)	0.569 (0.184)
Wooden crafts and instruments	0.268	-1.149 (0.253)	0.365 (0.181)	-0.615 (0.470)
Dyeing, cleaning and washing related worker	0.261	-3.418 (0.820)	1.116 (0.439)	1.265 (0.483)
Dress related workers	0.325	0.122 (0.542)	0.415 (0.305)	0.333 (0.441)
Leather workers	0.071	-1.647 (0.311)	0.529 (0.171)	0.718 (0.140)
Wood items related worker	0.408	-0.240 (0.199)	0.339 (0.130)	0.749 (0.080)
Metal related worker	0.450	-0.651 (0.228)	0.601 (0.148)	0.940 (0.076)
Glass, brick and porcelain related worker	0.417	-0.765 (0.225)	0.015 (0.190)	-0.673 (0.500)
Food and beverage producers	0.343	-1.132 (0.310)	0.782 (0.191)	0.796 (0.147)
Tobacco products	NA	-1.430 (0.175)	0.551 (0.131)	-1.588 (0.056)
Barbers and beauticians	0.475	-1.379 (0.238)	0.703 (0.150)	0.799 (0.108)
Construction	0.336	1.495 (0.040)	0.126 (0.073)	0.268 (0.024)
Workers in utilities (power, water, etc)	NA	-0.965 (0.221)	0.636 (0.137)	1.149 (0.066)
Printers, paper and book makers	NA	-3.492 (0.766)	1.543 (0.417)	1.521 (0.224)
Precision Instrument Makers and Repairers	NA	-1.583 (0.352)	0.990 (0.207)	1.245 (0.110)
Jewelers and Precision Metal Workers	0.448	-1.061 (0.170)	0.311 (0.112)	0.833 (0.059)
Garbage workers	0.078	-0.022 (0.531)	0.060 (0.353)	0.067 (0.504)
Transportation of all kinds	0.363	0.216 (0.088)	0.537 (0.080)	0.850 (0.035)
Post office, Telegraph and Telephone service	NA	-2.003 (0.436)	0.504 (0.245)	1.271 (0.127)
Financial intermediation	0.546	-3.976 (0.242)	2.180 (0.145)	1.786 (0.058)
Trade and retail shops	0.466	-0.451 (0.097)	1.168 (0.082)	1.110 (0.035)
Hotels	NA	-2.463 (0.353)	1.240 (0.208)	1.214 (0.135)
Music and entertainment	0.408	-3.898 (1.099)	1.474 (0.587)	1.632 (0.321)

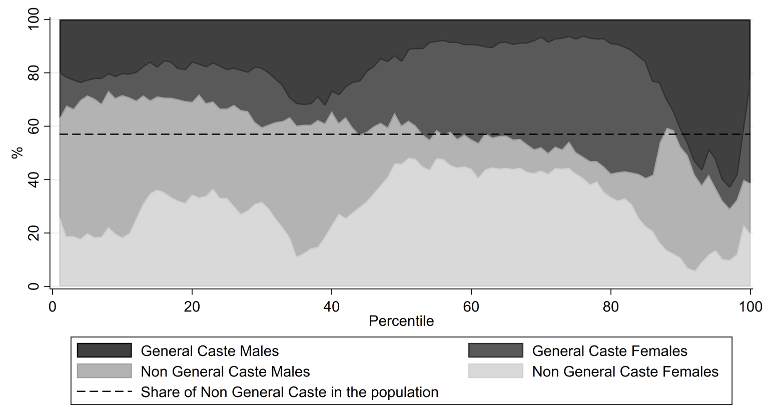
Table A9: Occupation-level Structural Parameters

	All-India occupation social rank	Occupation skill-wage ($\ln w_o$)	Occupation Amenity ($\ln A_o$)	Returns to skill ρ_o
	(1)	(2)	(3)	(4)
Protective services	0.715	-1.237 (0.142)	0.775 (0.097)	1.193 (0.047)
Government service	0.816	-3.072 (0.261)	1.053 (0.138)	1.671 (0.070)
Religious workers	0.883	-2.909 (0.512)	1.152 (0.311)	1.171 (0.142)
Legal professionals	0.750	-5.861 (0.808)	2.250 (0.397)	2.100 (0.171)
Doctors, modern and traditional	0.768	-3.171 (0.794)	0.881 (0.378)	1.588 (0.194)
Other medical professionals	NA	-4.943 (1.200)	2.085 (0.473)	2.176 (0.257)
Professors, teachers, education professionals	0.929	-7.655 (0.552)	3.400 (0.210)	2.890 (0.111)
Accountants, secretaries, clerks	0.750	-0.892 (0.162)	0.864 (0.098)	1.415 (0.041)
Architects, surveyors, engineers, and their employees.	NA	-3.657 (0.211)	1.052 (0.103)	1.983 (0.057)
High skill scientific or artistic	0.623	-2.078 (0.467)	0.808 (0.243)	1.271 (0.132)
Cooks and house servants	0.268	-0.256 (0.096)	0.065 (0.084)	-1.390 (0.061)
Manufacturers, business men and contractors otherwise unspecified	NA	-1.722 (0.214)	1.065 (0.125)	1.381 (0.064)
Mechanics otherwise unspecified	NA	-2.076 (0.264)	0.825 (0.159)	1.107 (0.086)
Home work	NA	0 (normalized)	0 (normalized)	0.747 (0.025)

Notes: Occupational social rank listed as “NA” for occupations that do not correspond to the traditional occupation of any caste.

A6 Appendix Figures

Figure A1: Distribution of human capital by gender and caste in baseline



Notes: This figure shows how men and women from either low or general castes are represented along the baseline human capital distribution. An example of the graph reading is: The bottom decile of the baseline human capital distribution is composed to 20 percent of general caste men, 10 percent of general caste women, 45 percent of non-general (i.e., low) caste men, and 25 percent of non-general (i.e., low) caste women.