Do others matter? : An empirical analysis of the interaction of social and human capital in India

Véronique Gille

Université Paris I Panthéon-Sorbonne
U.F.R. de Sciences Économiques

N° attribué par la bibliothèque

THÈSE

pour obtenir le grade de

Docteur de l'Université Paris I Panthéon Sorbonne

Discipline : **Sciences Économiques**

Titre:

Do others matter? An empirical analysis of the interaction of social and human capital in India

présentée et soutenue publiquement
par

Véronique Gille

le 2 décembre 2013

Directeur de thèse:
Jean-Claude Berthélemy

Jury

M.	Francis Bloch,	Professeur à l'Université Paris 1,	Président
Mme	Flore Gubert,	Chargée de recherche à l'IRD,	Rapporteur
M.	Kaivan Munshi,	Professeur à l'Université de Cambridge,	Rapporteur
Mme	Catherine Bros,	Maître de conférence à l'Université Paris-Est,	Suffragante
M.	Jean-Claude Berthélemy,	Professeur à l'Université Paris 1,	Directeur

L'UNIVERSITÉ PARIS 1 n'entend donner aucune approbation ou improbation aux opinions émises dans cette thèse. Ces opinions doivent être considérées comme propres à leur auteur.

Remerciements

Fallait-il vraiment faire une thèse pour répondre à la question « Do Others matter ? » ? Il m'aurait sans doute suffi de réfléchir à la liste des personnes que j'avais envie de remercier pour savoir que oui, others matter.

Je voudrais donc commencer par remercier mon directeur de thèse, Jean-Claude Berthélemy, qui m'a fait assez confiance pour m'accepter en thèse alors même que l'économie n'était pas ma matière principale, et qui m'a suivie pendant ces quatre années. Je voudrais aussi remercier les membres de mon jury, Francis Bloch et Flore Gubert, qui font désormais partie du groupe très restreint des personnes à avoir lu ma thèse en entier et Catherine qui tient plusieurs casquettes, puisqu'elle est à la fois membre de mon jury, ancienne collègue de bureau et mentor. Je dois énormément aussi à Kaivan Munshi, sans qui non seulement ma thèse, mais mon approche de la recherche en général ne seraient pas ce qu'ils sont.

Merci aussi à Elda et Loïc pour leur patience face à mes entrées intempestives dans leur bureau et à mes nombreuses questions administratives.

Qui dit thèse dit aussi doctorants. Et je n'aurais pas pu faire ma thèse sans le soutien (presque) journalier de ceux avec qui j'ai ou je partage encore le bureau. Un grand merci d'abord aux « vieux », ceux qui sont docteurs maintenant mais qui n'oublient pas ceux restés en arrière, Patrick, Yasser et Thomas, toujours là pour un bon conseil ou un bon mot. A Charlotte, pour son amitié. A ceux qui me précèdent de peu, Felipe pour toutes ses relectures, aides multiples et bonnes blagues, et Marie toujours là pour m'écouter et me remonter le moral, sans qui ces derniers mois auraient été moins faciles. A Diana, qui me suit depuis le master et avec laquelle j'ai partagé toutes les difficultés administratives liées à un début de thèse en janvier. A Ysaline de l'autre bâtiment. Aux « petits » qui ont bien grandi depuis, Victoire, Aurore, Francesco, Zorobabel, Daniel et Zaneta. Et enfin à ceux avec qui je partage depuis peu le luxueux bureau 314 (oui Manu tu seras cité ailleurs aussi), Anthony, Sophie, Manu, Margarita, Anna et Jamal.

Un grand merci à mes amis de « master ». D'abord les « recherches », qui ont été là à mes débuts quand je luttais encore pour comprendre une équation, et notamment Clément, Sébastien et Thomas, qui ont préféré l'argent à la science, Aurélien, qui m'a suivie dans cette aventure et Marion d'un soutien sans failles en toutes situations, qui

hoche la tête pour me rassurer pendant mes présentations et qui en plus d'être une vraie maman est une merveilleuse amie. Ensuite les « pros », Anda pour son soutien à ma recherche d'emploi, Manu dont les blagues se sont récemment améliorées, l'incontournable Esther et nos déjeuners sur la terrasse du 6ème, Adrien, Thaïs notre expatriée et Élise, ma fidèle amie, toujours là quand j'ai besoin d'elle (à Paris, Delhi ou Providence !), sans qui ma vie à Paris serait beaucoup moins drôle.

Je dois aussi des remerciements aux « indiens » et affiliés avec qui j'ai découvert ce pays qui nous est cher et qui fait l'objet de ma thèse, et qui m'ont bien fait comprendre que le quantitatif c'est bien, mais le qualitatif c'est mieux, Olivier, Khatidja, Julien, Thomas, Philippe, Clémence et Naïké.

Mes deux nouveaux colocataires, Lise et Matthieu pour avoir traversé ces deux derniers terribles mois avec moi.

Et enfin merci à tous mes autres amis, ceux qui ne comprennent pas très bien ce que je fais de mes journées, mais qui sont toujours là pour passer de bons moments. Les Sciences-Po, Anne-Lucie, Anne-Camille, Coralie, Sarah, Valentine, Marion, Tanguy, Édouard et Florent. Les Dijonnais, Anne-Sophie, Chloé et Juliette. Les inclassable, Marie, re-Marie, Elsa et Anne. Les Lilloises, Marine, Morgane, Eva, Emeline et Amandine. Sans oublier tous ceux rencontrés à Brown et particulièrement mes deux colocatrices, Yana et Paula.

Merci à Philipp, qui a beaucoup participé à cette thèse en rentrant lui aussi dans le club très sélectif des personnes à l'avoir lue en entier, et qui m'a appris que GMM ne voulait pas dire Arellano-Bond.

Et enfin un grand merci à ma famille, pour lesquels quelques mots ne suffiraient pas à exprimer toute ma gratitude d'être toujours là, chacun à sa manière, infailliblement.

Table of Contents

List of Tables

List of Figures

Résumé

Dire que le capital humain a un impact important sur le développement économique n'est pas une affirmation très controversée. Elle a été confirmée par la recherche des cinquante dernières années[1] et améliorer les indicateurs de capital humain dans les pays en développement est depuis quelques temps déjà un des objectifs principaux des gouvernements, des ONG et des organisations internationales. Cet engouement pour le capital humain se reflète bien dans les Objectifs de Développement du Millénaire qui ont pour but d'éradiquer la pauvreté à l'horizon 2015, quatre des huit objectifs énoncés visant directement ou indirectement à l'amélioration du capital humain[2].

Cependant, le capital humain est souvent seulement considéré comme une question individuelle, autant par les décideurs publics que par les chercheurs. Du côté politique, l'investissement dans le capital humain est vu comme un moyen d'améliorer les opportunités auxquelles les individus ont accès, et donc comme un outil efficace de lutte contre la pauvreté. Du côté de la recherche, cette perspective individuelle se reflète dans l'abondance de littérature sur les rendements privés du capital humain. Pourtant, le capital humain comporte une dimension sociale encore relativement peu analysée, bien qu'une littérature sur cette question soit en train d'émerger.

L'objectif de cette thèse est d'apporter un éclairage sur cet aspect social du capital humain. La question que je pose tout au long de cette thèse est « Quelle est l'importance des autres ? » par rapport au capital humain. En particulier, je me demande *comment capital social et capital humain interagissent*, et les éléments de réponse que j'apporte reposent sur des données indiennes. L'Inde est un pays où il est particulièrement intéressant d'étudier cette question, car le capital humain y a

[1] L'impact positif du capital humain sur le développement a été débattu (voir par exemple Benhabib et Spiegel, 1994 ; Pritchett, 2001), mais des articles récents montrent que l'absence de relation positive et significative entre éducation et croissance est liée à la qualité des données d'éducation (de la Fuente et Domènech, 2006 ; Cohen et Soto, 2007). De plus, il y a peu de doute dans la littérature sur le fait que l'éducation est importante pour d'autres enjeux de développement, et que d'autres composantes du capital humain (comme les compétences) ont un impact positif et significatif sur la croissance (voir par exemple Hanushek et Woessmann, 2008).

[2] Une description de ces objectifs est disponible sur le site des Nations-Unies http ://www.undp.org/mdg.

beaucoup évolué dans les cinquante dernières années, et le capital social a joué un rôle important dans cette évolution. Le contexte social indien, qui sera expliqué en détail dans la partie suivante de ce résumé, propose aussi un cadre d'analyse riche pour l'étude de la relation entre capital humain et capital social.

Avant d'aller plus loin dans ce résumé, je dois clarifier ce que j'entends par capital humain et capital social. Le capital humain est un terme global qui fait référence à la « capacité productive des êtres humains » (Schultz, 1961) et qui peut être amélioré par des investissements en « éducation, formation continue, santé, migration » (Becker, 1993). Cette liste est loin d'être exhaustive, et il y a beaucoup d'autres manières d'investir dans le capital humain, mais la littérature s'est surtout concentrée sur ces quatre composantes. L'approche adoptée pour cette thèse est plus restrictive, en ce qu'elle définit le capital humain par le niveau d'éducation des individus. C'est cependant une simplification assez courante[3], car l'éducation est la composante du capital humain qui a l'impact économique le plus direct (par exemple sur les salaires) mais aussi parce qu'elle est relativement facilement quantifiable.

Le capital social est aussi un concept complexe, et sa définition est toujours sujette à débats (Hayami, 2009). Il est néanmoins consensuel de considérer que le capital social se « concrétise dans les relations entre les personnes » (Coleman, 1988). En d'autres termes, selon Putnam (1995), le capital social peut être défini comme les « caractéristiques de la vie sociale - réseaux, normes et confiance - qui permettent aux participants d'agir plus efficacement en vue d'un but commun ». De nouveau ici je ne m'intéresse pas au capital social dans son ensemble, mais plutôt à sa dimension « réseau ». Dans cette thèse j'utilise la structure sociale indienne, le système de castes, pour mesurer le capital social auquel les individus ont accès. En effet, comme il sera expliqué dans la première partie de ce résumé, les interactions sociales en Inde (et particulièrement dans le monde rural) ont toujours lieu principalement au sein

[3]Becker (1993) lui-même dans son travail fondateur sur le capital humain restreint son analyse à l'éducation.

des castes.

Le reste de ce résumé est organisé de la manière suivante. La première partie explique le lien entre capital social et capital humain en Inde et la deuxième partie discute la littérature sur la relation entre capital humain et capital social. Enfin la troisième partie présente les différents chapitres de cette thèse.

Quel est le lien entre capital humain et capital social en Inde ?

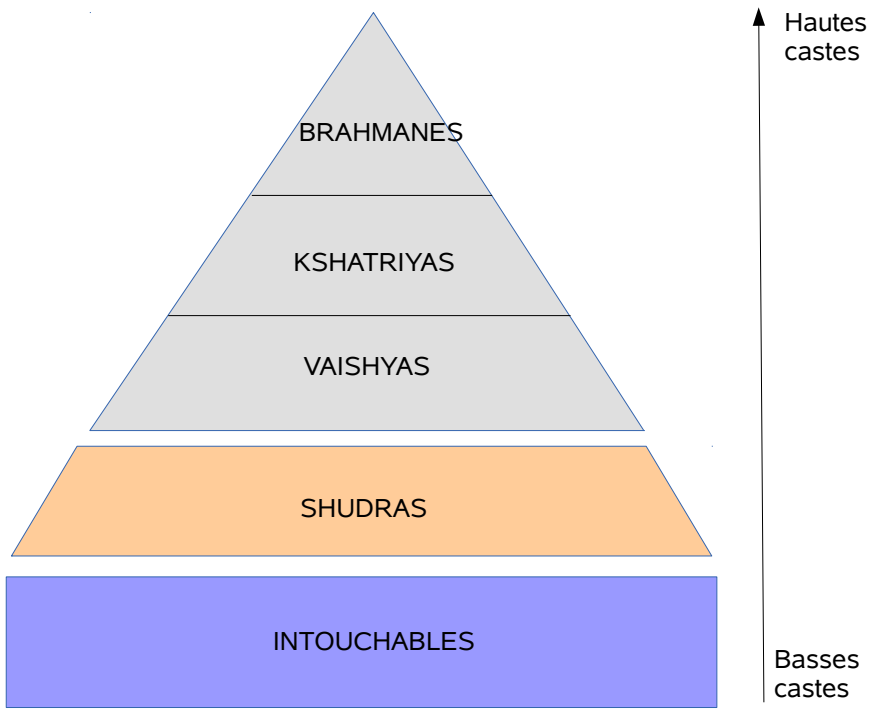
Le système de castes est un système très compliqué, avec beaucoup de règles, et pour bien le comprendre il faudrait bien plus que l'introduction que je vais proposer ici. Mais comme les interactions sociales en Inde sont très liées au système de castes, il est important d'avoir un aperçu de son fonctionnement pour comprendre cette thèse. Je commence donc dans une première section par expliquer les caractéristiques et les principales règles du système de castes avant de présenter dans la deuxième section le lien entre caste et capital humain en Inde. La position sociale dans le système de castes est un déterminant important du capital humain, mais comme souligné dans la troisième section cet état de fait n'est pas une impasse, car le système de castes permet, et est même parfois le vecteur, d'une certaine mobilité sociale.

Le système de castes

D'après la religion hindoue, la société indienne est divisée en quatre *varnas* que sont les Brahmanes (prêtres et professeurs), les Kshatriyas (princes et guerriers), les Vaishyas (commerçants) et les Shudras (serviteurs).

Le système de castes est souvent représenté par une pyramide comme dans la figure 1, où la position de chaque varna correspond à son rang dans la hiérarchie sociale indienne. Ainsi, les Brahmanes, Kshatriyas et Vaishyas sont considérés comme des hautes castes, alors que les Shudras sont considérés comme des basses castes.

FIGURE 1 – Le système de castes



Certaines personnes ne font partie d'aucun varna, ce sont les Intouchables, aussi appelés Dalits ou « Scheduled Castes »[4]. Les Intouchables sont exclus du système car ils sont considérés comme « impurs » du fait de leurs activités très dégradantes. Ils s'occupent par exemple des carcasses d'animaux, du nettoyage des sanitaires, etc. Les Intouchables sont par conséquent au plus bas de la hiérarchie sociale[5].

Cependant, concrètement la société indienne est divisée en une multitude de groupes, appelés *jatis*. Chaque jati appartient à un varna, et peut donc être situé dans la hiérarchie sociale globale. C'est autour du jati que s'organise la vie sociale. Chaque jati a une occupation traditionnelle, qui est souvent à l'origine de son nom. Par exemple, les Kurubas dans l'Etat du Karnataka sont des bergers, et leur nom

[4]L'origine de ce terme sera expliqué plus loin dans ce résumé.

[5]Savoir si les Intouchables ont historiquement été considérés comme impurs parce qu'ils faisaient des activités dégradantes ou s'ils ont été restreints aux activités dégradantes parce que la société les considérait comme impurs relève du débat de la poule et de l'œuf. Le fait est que la corrélation entre le type d'emploi et le degré de pureté est encore très forte aujourd'hui.

vient de « kuri » qui signifie mouton en langue kannada. L'occupation traditionnelle détermine aussi le degré de pureté ou d'impureté de chaque jati et ainsi son statut par rapport aux autres jatis. Les Intouchables, comme expliqué précédemment, ont traditionnellement des occupations très impures alors que les hautes castes ont des occupations beaucoup plus pures, puisqu'ils s'occupent de tâches plus intellectuelles comme l'enseignement ou la prière. Du fait de cette différence de pureté entre les différents jatis, les hautes castes peuvent être « polluées » si elles sont touchées par des basses castes ou si elles partagent la même nourriture ou la même eau. Pour préserver la pureté des jatis, les interactions et les échanges entre eux doivent donc suivre des règles spécifiques. Les Intouchables doivent par exemple marcher à une certaine distance des hautes castes et n'ont pas accès aux puits communautaires. Les principes d'hérédité et d'endogamie viennent compléter ce florilège de règles et de normes : le jati est hérité des parents donc l'appartenance à un jati est déterminée à la naissance et chacun doit se marier dans son jati. Il est donc très difficile d'échapper à son jati, et la mobilité sociale est très dépendante de la mobilité sociale du jati dans son ensemble.

Donc de quoi parle-t-on quand on parle de « caste » ? Ce terme est souvent utilisé pour parler du système dans son ensemble, c'est à dire à la fois du système de varnas et du système de jatis. J'utilise ce mot de la même manière dans cette thèse. Mais il y a des cas où j'utilise le terme de caste uniquement pour parler des varnas, comme dans le chapitre 2 ou uniquement des jatis, comme dans le chapitre 3 et 4. Je préciserai ce que j'entends pas caste quand le terme sera utilisé dans un sens plus spécifique.

Cette division de la société en castes, accompagnée d'une division du travail où les basses castes étaient restreintes à des tâches peu qualifiées, a créé une société très inégalitaire avec une répartition des richesses très corrélée à la position sociale de la caste. Bien qu'il n'y ait pas de contraintes réelles à ce que les basses castes s'occupent de tâches plus qualifiées, la discrimination dont elles souffrent a maintenu voire accentué cette situation au cours des siècles. Cependant, depuis l'Indépen-

dance la mobilité socio-économique a fortement augmenté, grâce à l'urbanisation et au développement économique de manière plus générale. Beaucoup d'individus ne travaillent plus dans l'occupation traditionnelle de leur caste, et la corrélation entre caste et situation économique n'est désormais plus parfaite[6]. Quelle est alors la pertinence de la caste aujourd'hui ? Peut-on vraiment soutenir que le capital social est créé principalement au sein de la caste ?

La réponse est oui. Bien que la société indienne ait beaucoup évolué ces dernières années, les chercheurs s'accordent sur le fait que le système de castes est toujours un déterminant fondamental des relations sociales dans l'Inde d'aujourd'hui. Certains chercheurs soutiennent même que l'identité de caste s'est renforcée au cours du XXème siècle. Dirks (2011) par exemple explique que le système de castes n'était pas aussi rigide avant l'arrivée des Britanniques, mais que ceux-ci ont participé à sa sécularisation. Ils ont notamment utilisé le système de castes pour asseoir leur domination, en attribuant des postes aux hautes castes dans l'administration coloniale. Dirks (2011) soutient aussi que la mise en place des recensements, qui référençaient la caste de chaque indien, a joué un rôle important. Selon lui, la verbalisation officielle de leur caste par les individus a créé un sentiment d'appartenance et une conscience de caste qui n'étaient pas aussi ancrés auparavant. Plus récemment, les politiques de discrimination positive, pour lesquelles le critère est la caste, ont eu le même impact. Ces politiques ont renforcé l'identité de ceux qui en bénéficient, mais aussi de ceux qui en sont exclus, parce qu'ils se sont constitués en groupes de pression soit pour s'opposer à cette politique soit au contraire pour essayer d'en bénéficier.

Deux exemples concrets illustrent le fait que la caste est encore aujourd'hui un déterminant important de la vie sociale. Premièrement, il y a eu ces dernières années une émergence de la politique de caste. De nombreux parties mobilisent leur électorat sur la base de la caste (Chandra, 2004). Deuxièmement, la règle de l'endogamie est encore très respectée, la norme étant toujours de se marier au sein

[6]Mais nous verrons plus loin dans ce résumé qu'elle est toujours forte.

de sa caste. De manière anecdotique, il est intéressant de voir que sur les sites de rencontre matrimoniaux, un des champs obligatoires est celui de la caste. Moins amusant mais tout aussi significatif de l'importance de la caste aujourd'hui, des histoires de meurtres de jeunes couples de castes différentes font souvent la une des journaux locaux, ceux-ci ayant été assassinés par leur famille pour s'être mariés en dehors de leur caste.

La caste est donc un système très contraignant, et sa persistance peut apparaître comme surprenante. Mais cela peut être en partie expliqué par le fait que c'est une institution qui peut aussi être utile à ses membres. Munshi et Rosenzweig (2009) montrent par exemple que le réseau de caste fonctionne comme un système d'assurance dans le monde rural. Par ailleurs, comme souligné précédemment, le système de castes a su évoluer pour être à même de saisir les nouvelles opportunités. C'est particulièrement vrai dans la sphère politique, où de nombreux groupes de castes sont se constitués en lobbys pour défendre leurs intérêts.

Quel lien entre castes et capital humain ?

La partie précédente souligne que la caste est une institution qui gouverne la vie sociale en Inde. C'est aussi un déterminant important du statut socio-économique des individus. Même s'il y a une une certaine convergence en termes de niveau d'éducation entre les différentes castes (Kijima, 2006 ; Desai et Kulkarni, 2008), le niveau d'éducation est toujours très corrélé à la position de la caste dans la hiérarchie sociale. Lanjouw et Stern (1998), qui ont étudié le village de Palanpur dans l'Uttar Pradesh pendant plus de cinquante ans, soulignent par exemple que dans ce village l'alphabétisation moyenne des castes est très corrélée à leur statut social dans le village. Ils rapportent même qu'« une analyse statistique suggère que la caste a un impact [sur le niveau d'éducation], qui est indépendant du revenu individuel (et même après avoir contrôlé pour le niveau d'éducation des parents) ». Cette tendance se confirme au niveau du pays. La caste n'est plus demandée dans

les enquêtes nationales depuis le recensement de 1931[7], donc il est impossible d'observer très précisément le lien entre niveau d'éducation et caste. Cependant dans les données il est possible de différencier les « Scheduled Castes » (SC), les « Scheduled Tribes »[8] (ST) et les « autres ». Et la différence entre les SC/ST et les autres est assez frappante : selon Kijima (2006)[9], en 1999 41,1% des ménages ST et 32,9% des ménages SC n'ont pas de membre qui sait lire, contre « seulement » 18,1% des ménages des autres groupes. De la même manière, 13,6% des « autres » ménages ont un membre qui a un diplôme universitaire ou supérieur, contre seulement 4,2% des ménages ST et 5,3% des ménages SC.

Cette situation est héritée de centaines d'années de discrimination à l'encontre des basses castes, qui n'avaient pas accès à l'éducation. Mais même aujourd'hui que les basses castes ont le droit d'aller à l'école, elles ne sont pas traitées de la même manière[10]. La première source d'inégalité tient au type d'école auquel les basses castes ont accès. Les écoles privées ont la réputation d'être meilleures que les écoles publiques, notamment dans le monde rural. Par conséquent, dans certains villages les hautes castes vont à l'école privée, tandis que les basses castes vont à l'école publique. Même quand les enfants de basses castes et les enfants de hautes castes fréquentent la même école, les hautes castes ont certains privilèges que les basses castes n'ont pas. Par exemple, dans certaines écoles les hautes castes peuvent s'asseoir sur des bancs, tandis que les enfants dalits doivent s'asseoir par terre. Des attitudes très discriminatoires des professeurs envers les basses castes, pouvant aller jusqu'aux sévices physiques, ont été rapportées (rapport PROBE, 1999). Hanna et Linden (2012) montrent aussi que les professeurs ont tendance à mettre de moins bonnes notes aux basses castes qu'aux hautes castes pour des réponses similaires. Ces

[7]Sauf dans le recensement de 2011, mais les données ne sont pas encore disponibles.

[8]L'appellation de Scheduled Tribes désigne les minorités ethniques. Ces minorités ethniques sont globalement très pauvres et peuvent donc bénéficier des politiques de discrimination positive.

[9]Les données proviennent de la 55ème vague du « National Sample Survey ». Les États du Nord-Est sont exclus.

[10]En 2011, le taux de scolarisation dans le primaire était de 100%. Les statistiques peuvent être trompeuses, d'autant plus qu'il est assez fréquent qu'un même enfant soit compté deux fois parce qu'il est inscrit à la fois à l'école publique et à l'école privée, mais ce pourcentage souligne malgré tout un accès presque universel à l'éducation.

discriminations ne sont pas restreintes au monde rural, ni à l'éducation primaire. Le rapport PROBE rapporte les propos d'un professeur à Delhi : « Pourquoi enseigner à des enfants Scheduled Castes ? Laissons-les apprendre à jouer du tambour, c'est déjà bien assez ». Dans certaines universités les dalits se plaignent de harcèlement, et cette situation a conduit certains au suicide[11].

Cette forte corrélation entre statut social de la caste et niveau d'éducation est renforcée par le peu d'opportunités dont bénéficient les basses castes. Leurs investissements dans l'éducation sont freinés par le fait qu'elles ont un plus faible rendement de l'éducation que les autres castes (Kijima, 2006).

Plusieurs raisons expliquent ce rendement plus faible de l'éducation. En premier lieu, dans la division du travail entre castes, les basses castes ont historiquement occupé des métiers demandant peu de qualifications. Bien que les choses changent, beaucoup de basses castes continuent à travailler dans le secteur traditionnellement rattaché à leur caste, et à occuper des activités où les rendements de l'éducation sont faibles. Munshi et Rosenzweig (2006) étudient un choc positif sur les rendements de l'éducation à Bombay, et montrent que les garçons de basses castes ne répondent pas à ce choc et n'investissent pas plus dans l'éducation qu'avant le choc. Au contraire les filles répondent positivement, en s'inscrivant notamment dans des écoles où l'éducation est en anglais. D'après eux, la différence de comportement entres filles et garçons est due au fait que les garçons sont liés au secteur traditionnel dans lequel ils doivent travailler alors qu'au contraire les filles ont la liberté de travailler en dehors de ce secteur, car elles n'y sont pas historiquement attachées[12].

Deuxièmement, travailler dans le secteur formel nécessite un réseau. Lanjow et Stern (1998) montrent que dans le village de Palanpur, les personnes qui ont un travail à l'extérieur du village ont tendance à avoir le même emploi ou à travailler

[11] Le film documentaire « Death of Merit » ? décrit cette situation. Il est disponible en ligne à http ://thedeathofmeritinindia.wordpress.com.

[12] Les filles ne sont pas liées au secteur traditionnel car très peu de femmes de la génération précédente travaillaient.

dans la même entreprise que les autres membres de leur sous-caste. Selon eux, cela « reflète la manière dont la recherche de travail est organisée dans ce secteur, qui passe plutôt par l'utilisation de 'contacts' plutôt que par une recherche 'impersonnelle' des futurs employés (ou employeurs). Ceux qui ont déjà un travail en dehors du village sont la plupart du temps bien placés pour aider leurs amis, leur famille ou les personnes de leur caste à obtenir les postes vacants dans l'endroit où ils travaillent ; et les employeurs utilisent souvent leurs employés comme recruteurs. » Il est donc plus compliqué pour les basses castes d'obtenir un emploi dans le secteur formel, car ils sont moins connectés à l'extérieur de leur village que les hautes castes. Et ça n'est pas seulement le cas pour le secteur privé. Nous verrons dans le chapitre 4 que bien que les basses castes bénéficient de quotas dans le secteur public, il est difficile pour eux de réussir à avoir un emploi dans ce secteur car le recrutement est principalement discrétionnaire et repose largement sur le réseau.

Enfin, il est n'est pas impossible que la discrimination dont souffrent les basses castes joue aussi un rôle dans ces plus faibles rendements de l'éducation. En effet, Iyer et al. (2013) soulignent que les hautes castes sont plus représentées dans le secteur privé que les basses castes, que ce soit parmi les dirigeants des entreprises ou parmi les employés.

Quelles possibilités pour les basses castes ?

Les deux sous-parties précédentes présentent un aperçu relativement triste de la situation des basses castes en Inde : elles semblent être enfermées dans une sorte de trappe à pauvreté, avec un faible niveau d'éducation et peu d'opportunités. Pourtant, dans une perspective dynamique la situation est beaucoup moins dramatique. Desai et Kulkarni (2008) montrent que l'écart entre les basses castes et les hautes castes en termes de niveau d'éducation et de revenu est en train de se réduire. Hnatkovska et al. (2013) soulignent aussi que la mobilité intergénérationelle des SC/ST a rejoint celle des non-SC/ST sur la période de 1983 à 2008. Ainsi les choses semblent être en train de changer. La section suivante explore deux sources de changement

pour les basses castes : l'utilisation de la force des réseaux de caste et les programmes de discrimination positive.

Le système de castes a été décrit comme un système empêchant la mobilité sociale. Pourtant, ça n'est pas toujours le cas. Damodaran (2008) décrit des trajectoires sociales de castes où les liens forts qui unissent les membres d'une même caste ont été exploités pour que les membres de la caste puissent profiter de nouvelles opportunités. Bien que dans le passé ces évolutions aient surtout été exploitées par des hautes castes, Munshi (2011) relate l'histoire des Kathiawaris, des basses castes historiquement travailleurs agricoles qui ont réussi à s'insérer dans l'industrie du diamant à Bombay en mutualisant leurs forces. La norme du mariage intra-caste a joué un rôle important car elle a permis de consolider les liens de castes dans un secteur où une grande confiance entre partenaires est nécessaire.

Mais ce sont surtout les politiques de discrimination positive qui sont les principaux leviers de mobilité sociale pour les basses castes. Les politiques de discrimination positive existent depuis longtemps en Inde puisque dès 1882 les Britanniques avaient créé des écoles pour les Intouchables (Jaffrelot, 2011). Mais c'est à l'Indépendance qu'elles ont vraiment pris leur essor. Malgré l'opposition des deux principaux leaders de l'Indépendence sur ce sujet[13], le principe de discrimination positive pour les Intouchables, maintenant appelés « Scheduled Tribes » (ST) a été inscrit dans la Constitution. Depuis cette époque, les SC et les minorités ethniques (les « Scheduled Castes », SC) bénéficient respectivement de 15 et 7,5% de quotas dans le secteur public, les universités publiques et pour les postes d'élus.

Cette politique de quotas s'est plus tard étendue à une autre catégorie de la population, les Shudras maintenant référencés sous le terme de « Other Backward Classes » (OBC). Bien que les OBC soient considérés comme des basses castes, la mise en place d'une politique de discrimination positive en leur faveur a été beaucoup

[13]Gandhi et Ambedkar avaient des idées contradictoires sur les moyens à mettre en œuvre pour améliorer le statut des Intouchables.

plus conflictuelle. En effet, les OBC constituent environ 50% de la population[14], donc leur faire bénéficier des quotas n'est pas anecdotique. De plus, contrairement aux SC et ST qui sont très majoritairement extrêmement défavorisés, la population des OBC est beaucoup plus hétérogène. Comme les OBC ne souffrent pas du stigmate d'intouchabilité, ils sont moins discriminés et certains ont pu améliorer leur situation en profitant des opportunités qui s'offraient. Certains jatis OBC dans leur ensemble ont aussi localement un statut social élevé du fait de leurs propriétés foncières.

Pour toutes ces raisons il a été compliqué de faire émerger un consensus national sur les quotas pour les OBC. L'État Central[15] a donc laissé une grande liberté aux États pour légiférer en cette matière. Les premiers États à établir des quotas pour les OBC similaires aux quotas pour les SC et les ST ont été les États du Sud (Kerala, Karnataka, Andhra Pradesh, Tamil Nadu) (Galanter, 1978), et les derniers les États de la ceinture Hindi (Rajasthan, Haryana, Uttar Pradesh, Madhya Pradesh). Finalement, en 1993 l'administration centrale a mis en place des quotas pour les OBC et depuis 2008 les universités nationales[16] doivent réserver 27% de leurs sièges aux OBC.

Au total, 49,5% des sièges dans les universités sont donc désormais réservés pour les basses castes. Ces politiques de quotas s'accompagnent d'autres aides plus en amont dans la scolarité, comme des repas gratuits dans le primaire, des bourses dans le secondaire, etc. Mais il n'y a pour l'instant pas de consensus sur l'efficacité de ces politiques de discrimination positive. Par exemple, Cassan (2011) utilise les changements de frontière des États pour évaluer l'impact des politiques de discrimination positive sur le niveau d'éducation des SC. Il ne trouve aucun impact, sauf dans les districts où l'offre d'éducation est abondante. A l'inverse Bertrand et al. (2010) montrent que les basses castes qui ont bénéficié des quotas dans les Universi-

[14]Cette estimation de la population OBC provient du rapport de la commission Mandal, qui avait pour mission de documenter le statut des OBC. Mais ce nombre est controversé, car il a été calculé à partir de chiffres du recensement de 1931 (le dernier recensement par castes).

[15]Le terme « Central » est ici utilisé pour tout ce qui décidé au niveau fédéral, et le terme « État » pour tout ce qui est décidé au niveau des États.

[16]En Inde il y a deux sortes d'universités : les universités nationales qui dépendent de l'État Central, et les université d'État qui sont administrées par les gouvernements des États.

tés s'en tirent mieux que ceux qui n'ont pas pu y accéder. Cependant, leur situation est toujours moins bonne que celle des hautes castes pour un même diplôme.

Quel est l'impact des autres ?

Dans la section précédente je me suis intéressée au lien entre capital humain et caste en Inde. Bien que le capital humain d'un individu en Inde soit extrêmement lié à la place de sa caste dans la hiérarchie sociale, il existe certains mécanismes qui permettent de rompre cette corrélation historique. Dans cette section, je continue à explorer le lien entre capital social et capital humain mais en prenant une perspective plus large, à la fois géographiquement (dans cette partie je m'intéresserai à d'autres pays que l'Inde), et sur le fond. Plus précisément, cette partie revient à la question initialement posée au début de ce résumé, *quel est l'impact des autres par rapport au capital humain ?* J'explore la littérature sur ce sujet en suivant deux perspectives, qui sont aussi celles suivies dans le reste de cette thèse : quel est l'impact du capital humain des autres sur la productivité ? et quel est l'impact des autres sur la formation du capital humain ?

Quel est l'impact du capital humain des autres sur la productivité ?

Ceux qui nous entourent ont un impact sur notre productivité et sur la productivité en général de trois manières. Le premier canal est celui des externalités de l'éducation. Si les travailleurs apprennent de leurs collègues plus éduqués, être entouré de personnes plus éduquées a un impact sur la productivité individuelle. Les deuxième et troisième canaux sont des effets de composition du capital humain. Le deuxième mécanisme dépend de la complémentarité entre les différents niveau d'éducation (Kremer, 1993). Si les tâches et les niveaux d'éducation sont complémentaires dans la fonction de production, la manière dont l'éducation est distribuée dans les entreprises ou dans l'économie dans son ensemble a un impact sur la productivité agrégée.

C'est aussi le cas pour le troisième canal. Si les rendements de l'éducation ne sont pas linéaires dans le niveau de capital humain (Elbers et Gunning, 2004), c'est-à-dire si le rendement marginal de l'éducation n'est pas le même pour des niveaux faibles et plus élevés d'éducation, alors la manière dont l'éducation est distribuée a aussi un impact sur la productivité agrégée.

Le premier canal a par conséquent un impact sur la productivité individuelle et sur la productivité agrégée, alors que l'impact du deuxième et du troisième canal ne peut se voir qu'au niveau agrégé. L'étude empirique de l'impact du capital humain des autres sur la productivité suit donc deux directions : une littérature étudie l'impact de la composition du capital humain dans l'économie sur la croissance ou la productivité au niveau agrégé, et dans ce cas il n'y a pas de distinction entre les trois canaux. Un deuxième pan de la littérature regarde directement l'impact des voisins ou des collègues sur la productivité individuelle, et s'intéresse donc seulement au premier canal, celui des externalités de l'éducation.

Au niveau agrégé, la littérature étudie principalement l'impact de l'inégalité dans l'éducation sur la productivité ou la croissance. C'est une littérature relativement récente, et dont les résultats sont pour l'instant contradictoires. Londono et Birdsall (1997), Lopez et al. (1998), Castelló et Domenech (2002) et Castelló-Climent (2010b) par exemple trouvent que l'inégalité de l'éducation a un impact négatif sur la croissance ou le revenu par habitant. Schipper et Hoogeveen (2005), Park (2006) et Rodríguez-Pose et Tselios (2009) en revanche trouvent un impact positif. Les raisons de ces désaccords sont étudiées en détail dans le chapitre 1. Mais il semble a priori clair que plus de recherche est nécessaire pour bien comprendre l'impact de la distribution de l'éducation sur l'économie d'un pays.

La littérature qui s'intéresse aux externalités de l'éducation est beaucoup plus développée[17]. Ici je ne mentionnerai que les articles qui me semblent les plus im-

[17]Moretti (2004b) fournit une revue de littérature sur cette question ainsi que sur l'impact du capital humain des autres sur d'autres questions

portants, soit à cause de leur stratégie d'identification, soit de par le contexte dans lequel sont étudiées les externalités.

Comme souligné précédemment, les externalités de l'éducation peuvent être étudiées au niveau agrégé. C'est ce que font Acemoglu et Angrist (1999), Ciccone et Peri (2006) et Moretti (2004a) qui comparent les salaires des travailleurs dans les villes avec plus ou moins de capital humain. Acemoglu et Angrist (1999) ne trouvent aucun impact, mais Ciccone et Peri (2006) et Moretti (2004a) observent qu'une part plus importante de travailleurs avec une éducation supérieure a un impact positif sur la productivité totale.

Les travaux qui étudient les externalités de l'éducation au niveau des entreprises trouvent aussi un impact positif de l'éducation des autres travailleurs. Martins et Jin (2010), qui utilisent des données de panel d'entreprises portugaises et Moretti (2004c) qui étudie des entreprises américaines trouvent d'importantes externalités de l'éducation. Les résultats sont moins clairs dans un contexte non-industriel. A ma connaissance, seulement trois articles s'intéressent à l'impact du capital humain environnant sur la productivité agricole. Appleton et Balihuta (1996) et Weir et Knight (2007) trouvent que l'éducation des voisins a un impact positif sur la productivité. Asadullah et Rahman (2009) n'observent aucun impact. De nouveau, les raisons pour lesquelles il n'y a pas de consensus sur cette question seront étudiées plus en détail dans le chapitre 2.

Parmi les articles cités, aucun ne s'intéresse spécifiquement à l'Inde. Dans la littérature au niveau agrégé, la plupart des travaux portent sur les États-Unis ou sur des pays développés. L'article de Lopez et al. (1998) est une exception, puisqu'il étudie l'impact de l'inégalité de l'éducation avec des données de panel sur 12 pays en développement entre 1970 et 1994. Au niveau microéconomique, la littérature qui étudie l'impact de l'éducation des autres travailleurs sur la productivité dans les entreprises repose uniquement sur des données de pays développés. En revanche les travaux sur le lien entre éducation des voisins et productivité agricole s'intéressent

uniquement aux pays en développement. Appleton et Balihuta (1996) et Weir et Knight (2007) étudient cette question dans des pays africains, et Asadullah et Rahman (2009) utilisent des données du Bangladesh. La seule littérature connexe qui s'intéresse à l'Inde étudie les effets de pairs dans l'adoption des nouvelles technologies dans l'agriculture. Deux articles sont particulièrement intéressants sur cette question : Foster et Rosenzweig (1995) trouvent qu'il y a des effets importants d'apprentissage dans l'adoption des graines à forte productivité. Munshi (2004) montre que l'effet d'apprentissage est plus faible quand la productivité dépend de caractéristiques individuelles inobservables.

Le fait que le niveau d'éducation des autres ait un impact sur la productivité au niveau agrégé ou au niveau individuel est un aspect de l'interaction entre capital humain et capital social. Mais une question peut-être encore plus important est comment capital humain et capital social interagissent *ex-ante*. Autrement dit, quel est l'impact du capital social sur la formation du capital humain ? La prochaine section donne un aperçu de la littérature sur cette question.

Quel est l'impact des autres sur la formation du capital humain ?

La question de l'impact des autres sur la formation du capital humain a été, historiquement, la première question étudiée dans la littérature sur le capital social. L'article fondateur de Coleman (1988), qui a introduit le concept de capital social s'intitule en effet : « Le capital social et la création du capital humain ». Cet article étudie l'impact du capital social dans la famille et dans la communauté sur la création de capital humain chez les enfants, et montre qu'il y a des corrélations importantes entre capital social et échec scolaire. Depuis ce premier travail, la recherche sur cette question a explosé. La perspective choisie est cependant assez restreinte : la plupart des articles de recherche se concentrent sur l'impact des performances

scolaires des autres enfants sur la performance scolaire, mesurée par les notes obtenues aux examens ou aux interrogations, ou le redoublement. Sur cette question, la recherche s'accorde pour dire que les performances des autres enfants ont un impact important (Sacerdote, 2011).

Une littérature peut-être plus en rapport avec les questions étudiées dans cette thèse questionne l'impact du voisinage sur la réussite scolaire. Il n'y a cependant pas de consensus dans ces travaux. Goux et Maurin (2007) par exemple trouvent qu'il y a un impact positif de la proportion de familles non-éduquées dans le voisinage sur la probabilité de redoubler des enfants. En revanche, Gibbons et al. (2013) montrent qu'un changement dans la composition du voisinage n'a pas d'impact sur les notes obtenues aux interrogations. Une des raisons avancées pour expliquer cette variation forte dans les résultats obtenus est que la relation entre effets de voisinage et réussite scolaire n'est pas linéaire. Patacchini et Zenou (2011) étudient par exemple le lien entre qualité du voisinage et éducation des enfants, et montrent que la qualité du voisinage et l'implication des parents sont complémentaires, et que l'impact des voisins est plus important pour les parents peu éduqués.

Plus globalement, il semblerait que l'entourage ait un impact sur les choix liés à l'éducation, et notamment sur les choix d'orientation. De Giorgi et al. (2009) par exemple montrent, en utilisant des données de l'Université de Bocconi, que le choix de matière principale est fortement influencé par le choix des camarades de classe, ce qui conduit à un mauvais appariement des compétences réelles des étudiants et des diplômes obtenus. De la même manière Marmaros et Sacerdote (2002) montrent que la probabilité d'avoir un métier très rémunérateur augmente avec le revenu moyen des parents des colocataires pendant la première année d'université[18]. Mais de nouveau la littérature n'est pas unanime sur la question. Arcidiacono et Nicholson (2005) observent ainsi que le choix de spécialité des étudiants en médecine aux États-Unis est peu influencé par les choix des pairs.

[18]Mais ils ne peuvent pas savoir si l'effet est réellement dû à l'influence de leurs pairs ou si c'est un effet de réseau.

Enfin, un pan important de la littérature questionne l'impact de l'entourage sur les opportunités, via l'impact de l'entourage sur l'emploi. Comme souligné précédemment, les opportunités ont un impact indirect sur la formation du capital humain, car les investissements en capital humain sont influencés par les rendements attendus du capital humain. Corcoran et al. (1980), Topa (2001) ou Wahba et Zenou (2005) montrent par exemple qu'une part importante des informations sur les emplois disponibles provient du réseau. Les membres du réseau servent aussi de référence pour trouver un emploi (Topa, 2011), et la durée du chômage est plus faible pour les personnes ayant un réseau plus important (Cingano et Rosalia, 2012).

Quel est le rapport de cette littérature avec l'Inde ? De nouveau, aucun des articles cités précédemment ne parle directement de l'Inde et très peu sont appliqués à des pays en développement. Cependant une littérature très liée est celle qui associe capital social et groupe ethnique. Une grande partie de ces travaux s'intéresse au rôle des réseaux ethniques sur l'emploi et la carrière (voir par exemple Munshi, 2003 ; Edin et al., 2003 ; Patacchini et Zenou, 2012). Un article particulièrement intéressant sur l'impact de l'ethnie sur l'investissement en éducation est celui de Munshi et Rosenzweig (2006), qui montre que les réseaux ethniques n'ont pas toujours un impact positif sur l'investissement en éducation. Dans cet article, ils soulignent que les réseaux de caste peuvent être inefficaces, car ils peuvent empêcher de profiter des nouvelles opportunités, et en l'occurrence ici d'une augmentation des rendements de l'éducation.

Description de la thèse

Cette thèse s'inscrit dans la littérature mentionnée dans la partie précédente, en étudiant l'impact du capital humain des autres sur la productivité agrégée et individuelle, et le rôle du capital social sur l'investissement dans l'éducation, que ce soit directement ou indirectement via les opportunités qu'il ouvre. La structure de

cette thèse suit globalement l'ordre choisi pour présenter la littérature dans la partie précédente. Les **chapitres 1 et 2** s'intéressent à l'impact du capital humain des autres sur la productivité, et les **chapitres 3 et 4** étudient l'impact de la caste sur les candidatures aux programmes de discrimination positive, qui facilitent l'accès des basses castes aux universités (chapitre 3) et aux emplois dans le secteur public (chapitre 4). L'ordre des chapitres peut sembler contre-intuitif au premier abord (on s'attendrait à commencer par la *création* du capital humain, pour finir par l'*impact* du capital humain), mais il découle en partie de la manière dont sont définis les « autres » dans chacun des chapitres. Au fur et à mesure, la définition des autres est de plus en plus restreinte. Dans le chapitre 1, la définition est très large, puisque les autres sont tous les autres travailleurs dans l'économie. Dans le chapitre 2 les autres sont les personnes de la même caste (caste au sens varna), et dans les chapitres 3 et 4, les autres sont seulement les membres du même jati. L'ordre des chapitres est aussi défini par le niveau d'agrégation des données. Le chapitre 1 a une perspective macroéconomique alors que les chapitres 2, 3 et 4 sont microéconomiques.

L'objectif du **chapitre 1** est de regarder l'impact du capital humain des autres sur la productivité agrégée d'un point de vue macroéconomique. Comme souligné précédemment, la littérature macroéconomique sur la question s'est surtout intéressée à l'impact de l'inégalité du capital humain. C'est aussi cette approche qui est choisie ici. Ce chapitre étudie la relation entre la distribution du capital humain et le revenu par habitant des États indiens en utilisant six vagues d'enquêtes ménage (les « National Sample Surveys ») entre 1983 et 2009-2010. Deux mesures de la distribution sont utilisées : un Gini du capital humain, et un indice de Theil comme test de robustesse. Comme le Gini du capital humain est structurellement fortement corrélé au niveau d'éducation moyen d'un État, et que cela peut poser des problèmes de colinéarité, j'utilise la méthodologie suggérée par Berthélemy (2006) pour séparer dans le Gini « l'effet niveau » de « l'effet composition ».

Ce chapitre montre qu'il y a une relation négative entre le revenu par habitant et

l'*égalité* de l'éducation. Ce résultat est robuste à l'utilisation d'un indice de Theil, et a peu de chance d'être dû à une migration sélective, où les personnes les plus éduquées migreraient vers les États les plus riches. De plus, ce chapitre apporte un éclairage sur la diversité de résultats dans la littérature en soulignant que la relation entre distribution de l'éducation et revenu par habitant n'est pas linéaire. La relation entre égalité de l'éducation et revenu par habitant est bien négative pour les États les plus riches, mais elle est positive (quoique pas toujours significative) pour les États les plus pauvres. Enfin ce chapitre fait une analyse empirique très préliminaire des canaux qui pourraient expliquer la relation estimée. Comme souligné précédemment, il y a trois canaux théoriques. La présence d'une relation entre égalité de l'éducation et revenu par habitant peut être due à des externalités, au fait que les rendements de l'éducation ne sont pas linéaires, et/ou que les travailleurs sont complémentaires dans leur niveau d'éducation. Les données pour étudier les canaux sont très limitées - elles ne couvrent pas toute la période - donc les résultats obtenus sont potentiellement fragiles. Cependant ils montrent qu'a priori les trois canaux jouent un rôle.

Ce chapitre souligne donc l'importance de prendre en compte dans les politiques éducatives le fait que la composition du capital humain est liée au revenu par habitant, et ce indépendamment du niveau moyen d'éducation. Cependant, bien que les résultats soulignent qu'à même niveau d'éducation moyen les États ayant une proportion de personnes avec une éducation supérieure plus importante ont un revenu par habitant plus important, la recommandation tirée de ce chapitre n'est pas d'arrêter d'investir dans le primaire pour seulement investir dans le supérieur. En effet, les personnes ayant un niveau d'éducation supérieur sont d'abord passées par le primaire et le secondaire, et il est donc nécessaire d'investir massivement dans ces deux niveaux d'éducation pour espérer emmener une proportion importante de la population à l'université. Par ailleurs, le fait que la relation entre distribution de l'éducation et revenu par habitant ne soit pas linéaire appelle à une certaine prudence dans le design des politiques éducatives. Une des recommandations politiques de ce chapitre serait plutôt de se concentrer sur les abandons après le primaire pour

favoriser la poursuite des études secondaires.

Le **chapitre 2** est dans la continuité directe du chapitre 1. Il étudie aussi le lien entre capital humain des autres et productivité, mais d'un point de vue microéconomique. Le contexte dans lequel cette question est étudiée est aussi plus restreint, puisque ce chapitre s'intéresse à l'Inde rurale. En effet, l'objet d'étude est la productivité agricole : le but est d'étudier l'existence d'externalités de l'éducation dans l'agriculture en Inde, en regardant si le niveau d'éducation des voisins a un impact sur la productivité agricole. Un des problèmes majeurs dans l'analyse des effets de pairs est celui de l'endogénéité de la formation du réseau. Afin d'éviter ce problème, pour définir le groupe de référence, j'utilise le fait qu'en Inde rurale, comme souligné précédemment dans ce résumé, la plupart des interactions ont lieu au sein de la caste. Les données utilisées sont des données d'enquête ménage, provenant de la base ARIS-REDS, une base représentative de l'Inde rurale.

Les résultats soulignent qu'il y a des externalités de l'éducation : l'augmentation d'une année de l'éducation moyenne des membres de la même caste augmente la productivité agricole des ménages de 5%. De plus l'augmentation d'une année de l'éducation moyenne des membres du même village, mais pas de la même caste, n'a aucun impact sur la productivité agricole. De la même manière, les personnes de la même caste, mais qui ne cultivent pas la terre, n'ont pas d'impact. Ces tests permettent d'écarter l'hypothèse que les résultats obtenus sont des corrélations fallacieuses dues à des caractéristiques inobservées. En effet, pour obtenir les mêmes résultats sans qu'il n'y ait d'externalités, il faudrait que les caractéristiques inobservées soient totalement non-corrélées entre castes à l'intérieur du village et entre occupations au sein d'une même caste[19]. Ce chapitre met aussi en avant d'autres résultats, et notamment le fait que l'éducation des personnes de la même caste et l'éducation du chef de famille sont substituables : l'impact de l'éducation des per-

[19]Ce point a été mis en avant par Munshi et Myaux et est expliqué plus en détail dans le chapitre 2.

sonnes de la même caste décroit avec le niveau d'éducation du chef de famille. Les externalités sont aussi plus faibles pour le riz que pour le blé. Ce résultat est cohérent avec les résultats de Munshi (2004), qui montre que l'influence des voisins dans l'adoption de graines à haute productivité est plus importante en Inde pour le blé que pour le riz, car le transfert d'informations sur des nouvelles technologies est plus important quand la productivité de la technologie ne dépend pas de caractéristiques inobservables.

Ces résultats soulignent donc l'importance de continuer à investir dans l'éducation. En effet, l'éducation a non seulement des rendements privés, mais aussi un effet multiplicatif puisque l'éducation d'un individu bénéficie aussi à ses voisins. Du fait de la présence d'externalités cet investissement doit être financé publiquement. Par ailleurs, ce chapitre souligne que les externalités sont uniquement intra-goupes, et donc que l'éducation d'un groupe ne bénéficie qu'aux membres de ce groupe. Pour un impact optimal sur la productivité, il est donc important d'augmenter le niveau moyen d'éducation dans chaque groupe. Ce résultat rappelle la nécessité d'assurer l'égalité des chances des enfants quelle que soit leur caste, ce qui est encore loin d'être le cas en Inde aujourd'hui. Les politiques de discrimination positive semblent cependant répondre, dans une certaine mesure, à cet objectif.

Après avoir montré que le niveau d'éducation des autres a un impact sur la productivité, cette thèse s'intéresse à l'impact des autres sur la formation du capital humain. Dans le **chapitre 3**, j'étudie les déterminants à la (non-) candidature aux programmes de discrimination positive dans l'éducation en Inde, en me concentrant sur le rôle du stigmate, le stigmate étant défini comme la désutilité provenant de la participation à un programme. Comme expliqué précédemment dans ce résumé, l'Inde a mis en place des programmes de discrimination positive massifs en faveur des basses castes depuis l'Indépendance. Cependant, cette politique est fortement stigmatisée et les basses castes ayant bénéficié de quotas dans les universités sont victimes de discriminations et de harcèlement. Ce stigmate peut donc être un frein

à la candidature aux quotas. Pour étudier cette question, je me concentre sur les OBC (Other Backward Classes) et j'étudie l'impact du statut social du jati sur la probabilité de candidater aux quotas dans les universités. J'exploite le fait que le statut social d'un jati peut varier d'un village à l'autre, et qu'il est très corrélé à la proportion de terre possédée par ce jati dans le village. Les données utilisées sont les mêmes que dans le chapitre 2, mais je me concentre sur les États ayant mis en place des politiques de discrimination positive en faveur des OBC depuis au moins 10 ans.

Les résultats montrent que les membres de jatis ayant un statut social plus élevé ont moins tendance à postuler aux quotas, même après contrôle de la richesse individuelle. Cet effet augmente avec la richesse individuelle : les personnes les plus riches sont les plus affectées par le statut social de leur jati. Ce résultat est cohérent avec un effet de stigmate et je montre dans ce chapitre que les explications alternatives au stigmate ne sont pas soutenues par certains tests.

Ce chapitre montre que le faible taux de participation observé dans la plupart des programmes d'aide sociale peut probablement être expliqué en partie par un effet de stigmate. Mais bien que dans le cas d'autres programmes le stigmate puisse avoir des conséquences dramatiques, ça n'est pas le cas dans ce contexte précis. En effet, le taux de participation des SC/ST n'est pas affecté et parmi les OBC ce sont les plus riches parmi les jatis les plus puissants qui sont les plus touchés. Mais il faut garder à l'esprit que même si le stigmate n'a pas d'impact sur la candidature des plus défavorisés, il peut tout même affecter leur bien-être. Ce chapitre souligne donc l'importance d'une réflexion sur les solutions possibles pour dé-stigmatiser les quotas.

Enfin, le **chapitre 4** s'intéresse aux quotas pour les basses castes dans l'administration publique. En permettant aux basses castes d'accéder à des emplois qui étaient par le passé principalement détenus par des hautes castes, cette politique de quotas affecte leur rendement de l'éducation. Cependant, dans ce chapitre l'objectif

n'est pas d'estimer l'impact de ces quotas sur les investissements des basses castes en éducation (bien que ça soit une question intéressante). J'étudie la question des quotas sous un angle différent, mais lié. Je m'intéresse à *l'accès effectif* des basses castes aux quotas. En effet, pour que les quotas aient un impact sur l'investissement en capital humain des basses castes, il faut que les basses castes puissent réellement obtenir les postes qui leur sont réservés dans l'administration. Et ça n'est pas une question évidente dans le cas de l'administration indienne. Chandra (2004) souligne que pour plus de 95% des emplois dans le secteur public, le recrutement est discrétionnaire. Ce qui veut dire que les employeurs ont une grande liberté pour choisir leurs employés, mais aussi qu'il est difficile d'accéder à ces emplois sans avoir de connections.

Dans ce chapitre j'étudie l'impact du réseau de caste sur l'accès aux postes soumis à quotas dans l'administration publique. De manière générale, les individus ont tendance à s'appuyer sur les intermédiaires et les réseaux pour accéder aux bénéfices étatiques (Witsoe, 2012). Comme les données ne permettent pas de savoir si les personnes ayant candidaté aux quotas ont effectivement accédé à un poste, j'étudie si le fait d'avoir un réseau de caste mieux connecté à l'État a un impact sur le fait de *candidater*. Plus concrètement, j'analyse l'impact sur la probabilité de postuler aux quotas dans le secteur public, du fait d'avoir un élu local de la même caste. Les élus locaux en Inde ne sont pas directement en mesure d'attribuer les emplois dans l'administration. Mais ils ont un certain pouvoir sur les bureaucrates, car ils peuvent les transférer d'un service à l'autre (Chandra, 2004; Iyer et al., 2012). Je m'intéresse donc ici à un mécanisme indirect. Les élus locaux, de par leurs connections peuvent jouer un rôle d'intermédiaire entre les membres de leur caste et les bureaucrates.

Pour étudier cette question j'utilise des données issues de la même base que dans les chapitres 2 et 3, la base ARIS-REDS, qui fournit les informations nécessaires pour 37 villages dans trois États d'Inde du Sud. La stratégie d'identification est basée sur la réforme institutionnelle de 1993. A partir de cette réforme, lors des

élections de conseils de village, le poste de président du conseil est devenu réservé aux basses castes dans un certain nombre de villages par rotation. Les résultats montrent qu'avoir un président du conseil de village de la même caste augmente la probabilité de postuler aux quotas dans le secteur public. En explorant les canaux qui peuvent conduire à cet impact, je trouve que c'est l'effet de patronage qui correspond le plus aux résultats. Autrement dit, le président semble utiliser ses connections politiques pour aider les membres de sa caste à trouver un travail dans le secteur public. D'autres mécanismes, tels qu'un meilleur accès à l'information ou une augmentation de la confiance en soi ne semblent pas être prédominants d'après les tests présentés.

Ces résultats montrent que l'accès aux quotas n'est pas le même d'une caste à l'autre, et que les castes les mieux connectées sont celles qui bénéficient le plus des politiques de discrimination positive. Les rendements de cette politique, et par conséquent les investissements en capital humain dépendent ainsi des caractéristiques initiales des castes. Dans ce contexte particulier où le recrutement dans la fonction publique est principalement discrétionnaire, l'impact des politiques de discrimination positive semble donc rentrer en conflit avec l'objectif de départ affiché, qui était de favoriser l'accès des personnes les plus démunies aux emplois du secteur public.

Cette thèse, en étudiant l'interaction entre capital humain et castes en Inde, contribue à la littérature émergente sur l'impact du groupe sur les comportements individuels. Elle met notamment en avant le fait que le capital humain est relié de différentes manières au capital social en général et à la caste en particulier, et que la caste est toujours au cœur des institutions sociales en Inde.

Cette thèse apporte aussi un éclairage sur deux questions actuellement au cœur des débats en Inde. Les chapitres 1 et 2 contribuent au débat sur le rôle croissant de l'éducation privée en Inde. Comme déjà mentionné précédemment, ces chapitres, en soulignant la présence d'importantes externalités de l'éducation, rappellent l'importance d'une politique éducative cohérente, qui ne repose pas uniquement sur les

écoles privées.

Les chapitres 3 et 4, qui montrent que le fait de candidater à la discrimination positive dans les universités et dans l'administration publique dépend des caractéristiques de la caste et notamment de son pouvoir social, soulignent la nécessité d'une évaluation plus globale des politiques de discrimination positive basées sur des critères de caste.

Introduction

There is nothing controversial in saying that human capital matters a great deal for economic development. Research during the past 50 years has confirmed this belief,[20] and governments, international organizations and NGOs have worked hard to improve human capital indicators. The Millennium Development Goals, which aimed at eradicating poverty by 2015, are a good example of this increased interest: four out of the eight goals are directly or indirectly concerned with human capital.[21]

The majority of policy makers and researchers have considered and studied human capital as an issue only concerning individuals. On the policy side, investing in human capital is seen as an effective way of decreasing poverty because it opens up new opportunities to individuals. On the research side, the huge literature on the private returns to human capital is an example of this focus. However, human capital also has a social component which has not yet been well understood, despite a growing literature looking beyond the individual aspect of human capital.

The aim of this dissertation is to shed some light on this social component of human capital. The recurrent question that I am asking throughout this thesis is "*How do others matter?*", in relation to human capital. In particular, I am wondering *how social capital interacts with human capital*. To study this question, I take India as a case study. India is a country where human capital has dramatically changed in the last 50 years, and social capital had an important role in this evolution. More concretely, India's peculiar social structure, that is explained in details in the following section, provides a very interesting context to study the relation between human capital and social capital.

Before going further into this introduction, I need to clarify what I mean by human capital and social capital. Human capital is a global term which refers to "the productive capacity of human beings" (Schultz, 1961) and can be increased through investment in "schooling, on-the-job training, medical care, migration" (Becker, 1993). This list is far from being complete, and they are many other ways to invest in human capital, but the literature has mainly focused on these four components of human capital. In this dissertation, the focus is even more restrictive:

[20]The positive impact of human capital on growth has been debated (see for example Benhabib and Spiegel, 1994; Pritchett, 2001), but recent evidence shows that the absence of positive and significant relation between education and growth is related to the quality of educational attainment data (de la Fuente and Domènech, 2006; Cohen and Soto, 2007). Moreover, there is little doubt in the literature that education improves other outcomes related to development, and that other components of human capital (such as skills) have a positive and significant impact on growth (see for example Hanushek and Woessmann, 2008).

[21]A description of the goals is available on the United Nations website http://www.undp.org/mdg.

human capital is defined as people's education level. This is however a common simplification,[22] notably because education is the component of human capital which has the most direct economic impact (for example on wages) but also because it is is easily quantifiable.

Social capital, too, is a complex concept, and there is still a lively debate on how to define it (Hayami, 2009). However, there is a consensus on the fact that social capital is "embodied in relations among persons" (Coleman, 1988). In other words, following Putnam (1995), social capital can be defined as "features of social life - networks, norms, and trust- that enable participants to act together more effectively to pursue shared objectives". Again here, my focus is not on social capital as a whole, but rather on the network component of social capital. More precisely, in this dissertation, I use the social structure of India, the caste system, to define the social capital to which individuals have access to. Indeed, as it is explained in the first section of this introduction, interactions in India (and especially in rural India) are still mainly happening within castes.

The rest of this introduction is organized as follows. Section 1 first discusses how castes and human capital are related in India, before section 2 provides an overview of the literature on the relation between human capital and social capital. Finally, section 3 presents the outline of this dissertation.

1 How do social capital and human capital interact in India?

The caste system is a complicated institution with lots of rules, and getting a clear picture of how it works would require more than this introduction. However, social interactions in India (and therefore this dissertation) cannot be understood without basic knowledge on castes. Therefore, in this section I first explain the essential features of the caste system and the rules which regulate its functioning,[23] before focusing on how castes relate to human capital in India. We will see that the hierarchy in the caste system is a strong predictor of human capital, but this situation is not a dead end, as it is explained in the third subsection, because there is room for castes' mobility.

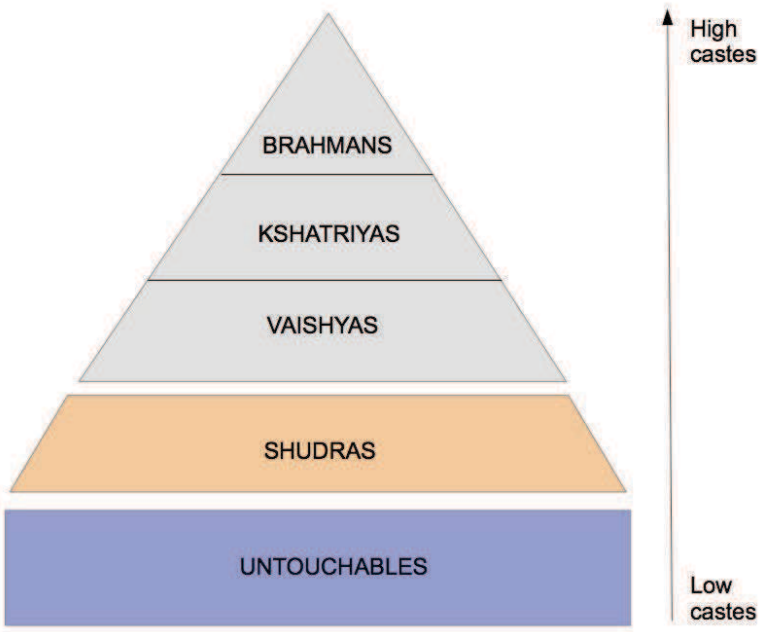
[22] Becker (1993) himself in his seminal work on human capital actually restricts his analysis to education.

[23] For readers who want to know more about castes, a more complete description of the system can be found in Bros (2010).

1.1 The caste system

According to the Hindu religion, the Indian society is divided into four *Varnas*: the Brahmans (priests and teachers), the Kshatriyas (rulers and warriors), the Vaishyas (traders) and the Shudras (servants).

Figure 2: The caste system



The system is often represented as in figure 2, where the position of each Varna in the pyramid represents their hierarchical position. More precisely, Brahmans, Kshatriyas and Vaishyas are considered as high castes. Shudras are low castes, and are hierarchically lower than all the others. Apart from these four groups, there are also people who do not belong to any Varna because their have occupations considered as profane or dirty, such as dust cleaning or scavenging. They are called the Untouchables or Dalits (or Scheduled Castes, this term will be explained later). They are hierarchically at the very bottom of the society.

In practice however, the Indian society is divided in thousands of groups called *jatis*. Each jati relates to a specific Varna and can consequently be situated in the global hierarchy. The jati is the real group of reference in India, along which social life is organized. Each jati has a traditional occupation which often gave its name to the group. For example, the Kurubas in Karnataka are traditionally shepherds, and they derive their name from "Kuri" which means "sheep" in Kannada. The traditional occupation also determines the degree of *purity* or *impurity* of each jati, which in turn determines their status compared to the other jatis. Untouchables have very polluting occupations, whereas high castes have purer occupations, more related

to intellectual activities, such as teaching or religious activities. Consequently, high castes can be "polluted" if they are touched by low castes for example or if they share the same food. Interactions and exchanges between jatis have therefore to follow specific rules so that purity can be preserved. Relations between jatis are also governed by other rules, like those of *heredity* and *endogamy*. The jati one belongs to is the same as the one of her/his parents, and one has to marry within her/his jati. It is therefore hard to escape one's jati, and social mobility is strongly dependent on the mobility of the jati as a whole.

So what do we mean when we talk about castes? The term "castes" is often used to refer to the system as a whole, the Varna and the jati system. In this dissertation, the term caste is used in the same way. However, there are some situations where the term caste refers specifically to the Varnas, as in chapter 2 for example, or to jatis, as in chapter 3 and 4. It will be made explicit when it is the case.

One consequence of the caste system, and in particular of the division of labor across castes which restricted in the past low castes to low-skilled jobs, is the creation of a very unequal society, where wealth is divided along caste lines. This situation has been maintained by the discrimination that low castes were and still are facing. However, since Independence, individual socio-economic mobility has increased thanks to urbanization and development, and households often have another activity than the traditional one of their jati. The correlation between caste and economic status is therefore not anymore perfect.[24] So what is the relevance of castes today? Does it make sense to consider that social capital is mainly created within castes in India?

The answer is yes. Scholars agree to say that, although the system has evolved a lot in recent years, caste is still what defines social relations in contemporary India. Indeed, caste identity is very strong and some researchers even claim that it has been strengthened during the 20th century. Dirks (2011) for example argues that the caste system before colonization was not as rigid, and that the British participated in the reinforcement of caste identity because they used the caste system to establish their authority. The censuses that they conducted also played an important role, because people were asked about their caste. Dirks affirms that the official verbalization of people's identity created a sense of belonging to their caste group. Caste identity has also been accentuated by the affirmative action policies put into place after Independence, for which the eligibility criteria is caste-based. Not only the caste identity of those being eligible has been reinforced, but also the identity of those not eligible, who gathered to oppose the policy, or who constituted themselves into

[24] Although we see in section 1.2 that the correlation is still strong.

pressure groups to get access to the same benefits.

Two examples can illustrate the fact that castes are a very important determinant of social life today. The first example is the development of caste politics: parties where mobilization is based on castes have emerged all over India (Chandra, 2004). The second example is that the rule of endogamy is still enforced. Marriage within the subcaste is still the norm, and there is also anecdotal evidence that intra-caste marriage does not only concern the old generations. For example, matrimonial websites are flourishing in India, and for most of them, declaring the jati is a pre-requisite. Stories about young couples who were killed by their families because they were not from the same caste are also widely spread.

The caste system is very constraining, and its persistence can seem surprising. But one important reason why caste still exists today is that it is also a useful institution for its members. Munshi and Rosenzweig (2009), for example, show that castes serve as insurance networks in rural India. While this an example of a role inherited form the past, castes have also evolved to take advantage of new opportunities. This is particularly true in the political sphere, where as previously mentioned, some caste groups act as lobbies to defend their groups' interests.

1.2 How are castes related to human capital?

The previous section highlighted the importance of caste as an institution governing social life in contemporary India. In parallel, caste is also a strong determinant of socio-economic outcomes and in particular of educational attainment. Even if some convergence in educational attainment across caste groups has occurred (Kijima, 2006; Desai and Kulkarni, 2008), there is still a clear relation between the hierarchical position of a given caste group and its education level. Lanjouw and Stern (1998), who have been studying the village of Palanpur in Uttar Pradesh over five decades, note that in this village caste literacy rate is correlated to the caste status and even that "statistical analysis suggests that caste has an influence of its own [on the education level], independently of per capita income (and even after controlling for parental education)". This pattern is confirmed at the national level. Since 1931, there has been no official Indian survey where people were asked for their caste group,[25] so it is impossible to exactly match education level and caste. However, data provide separate information for "scheduled castes" (SC), "scheduled tribes" (ST) and "others". And the gap between SC/ST and the others is striking: according to Kijima (2006),[26] in 1999, 41.1% of the ST and 32.9% of the SC households

[25]Except in the Census of 2011, but data are not yet available.

[26]Data are from the 55th round of the national sample survey. North-eastern States are not included.

had no literate member, against "only" 18.1% of the other groups. Similarly, 13.6% of the other households had a member with a university level or above, against only 4.2% of the ST households and 5.3% of the SC households.

This situation has been inherited from centuries of discrimination against low castes, who had no access to education. But even now that low castes have equal access to education,[27] they do not have equal treatment. The first source of inequality is through the kind of school that low castes can access. Because private schools tend to be of better quality than public schools in India, in some villages a situation has emerged where low castes students go to public schools and high castes students go to private schools. And even when low castes students attend the same schools as high caste students, they still suffer from differential treatments: for example dalits have to sit on the floor when their counterparts seat on benches. Discrimination does not only come from other students. It is also reported that teachers have negative attitudes towards low castes students, going from indifference to physical punishments [PROBE report, 1999]. Hanna and Linden (2012) also find that teachers give lower grades to low castes than to high castes for similar answers in exams. Moreover, these phenomena are not restricted to rural areas. The PROBE report quotes a teacher in Delhi: "What is the point of teaching scheduled-caste children? Let them learn how to beat drums, that's good enough.". Low castes students also suffer from discrimination at higher level of education. Harassment against dalits by other students has been reported in elite universities, and in several cases it even conducted them to commit suicide.[28]

This strong correlation between caste status and educational attainment is reinforced by the lack of opportunities that low castes face. Indeed, low castes have a lower return to education (Kijima, 2006) and have therefore less incentives to invest in education.

There are several reasons for these lower returns to education. First, it is related to the division of labor across castes, and to the fact that low castes were historically restricted to low-skilled jobs. Even though things are changing, most low castes continue to work in the area where members from their jati work, where returns to education are low. Studying the response of low castes to a positive shock to returns to education in Mumbai, Munshi and Rosenzweig (2006) find that the answer is stronger for low caste girls than low caste boys, because boys are expected to

[27] In 2011 school enrollment rate in India was 100%. The numbers can be misleading, because of the double counting of children who are enrolled in public and private schools, but it really denotes an almost universal access to education.

[28] The documentary film "Death of Merit"? describes this situation. It is available online at http://thedeathofmeritinindia.wordpress.com/.

work in the subcaste traditional occupation, which does not require high education levels. On the contrary girls are more responsive in terms of educational investment, because they were historically kept away from the labor market and are therefore not tied up to the traditional occupation.

Second, getting employment in the formal sector requires connections. Lanjouw and Stern (1998), looking at the outside jobs secured by the inhabitants of Palanpur, find that all the members from a subcaste having a job outside of the village tend to have the same job or work in the same company. Their interpretation of this pattern is that it "reflects the nature of the job search process in this segment of the labour market, which operates through 'contacts' rather than through 'impersonal' search by prospective employees (or employers). Those who have already secured a job outside the village are usually in a privileged position to help their friends, relatives, or fellow caste members to take advantage of possible vacancies in their own place of employment; and employers themselves often use their existing employees as recruiting agents". Therefore, it is more complicated for low castes to secure white-collar jobs, because their network outside of the village is less developed than the one of higher castes. And this is not only true for private sector jobs. Although there are quotas for low castes for jobs in the public sector, connections are also needed because the hiring process is highly discretionary. This point is further explained in chapter 4.

Finally, it is possible that there is discrimination against low castes in the recruitment process. Indeed, high castes are overrepresented in the private sector, in the ownership of firms as well as in the workforce (Iyer et al., 2013).

1.3 Improving low castes' human capital

The two previous subsections give a very negative overview of the situation. Low castes seem to be stuck in a poverty trap, with low education levels and little opportunities. However, according to the literature, the educational gap between low castes and high castes seems to be slowly shrinking (Desai and Kulkarni, 2008). Moreover, intergenerational mobility of SC/ST in terms of education and income has reached the level of non SC/ST during the 1983-2008 period (Hnatkovska et al., 2013). These facts underline that things can and are actually changing. This section focuses on two different possibilities for low castes to improve their economic position: the use of caste networks' strength and affirmative action programs.

The caste system has been described as an institution which prevents mobility. However, there are cases where the strong links that connect castes members to each other have been used by castes groups to move into new occupations. Although most

of the success stories are of high castes (Damodaran, 2008), Munshi (2011) describes the case of a low caste of agricultural laborers, the Kathiawaris, who succeeded to enter the diamond industry business in Mumbay. The rule of intra-caste marriage was notably used as a tool to reduce commitment problems.

However, the principal instruments available to low castes to improve their economic status are affirmative action programs. Affirmative action in India has a long history: as early as 1882 the British created special schools for Untouchables (Jaffrelot, 2011). But it was after Independence that reservation policies were implemented on a large scale. Although there was no consensus between Independence leaders on how to improve low castes status,[29] "reservations" in favor of the Untouchables, now called the "Scheduled Castes" (SC) were written in the Constitution. The SC, as well as the ethnic minorities (called Scheduled Tribes, ST) were provided with respectively 15% and 7.5% quotas in the administration, in public universities and in elections.

This reservation system was latter extended to the Shudras. The extension of the system of quotas to this category of population, now called the "Other Backward Classes" (hereafter OBC) was however very gradual and a large freedom was left to States governments to legislate on this matter. The reason for this flexibility on reservations for OBC was due to the controversial aspect of this policy. The OBC constitute almost 50% of the Indian population.[30] Therefore, extending the quotas to this whole population is not anecdotal and lowers the number of remaining seats for the other castes. The OBC population is also very heterogeneous, contrary to the SC/ST who are almost without any exception very deprived. Given that they suffer less from discrimination, prior to reservations, some OBC had the opportunity to improve their economic status and cannot be considered anymore as disadvantaged. Moreover, among the OBC, some jatis as a whole enjoy locally very influential positions due to their landholding position.

This situation made it complicated for the Central[31] government to come up with a consensus on reservations for OBC. Therefore, the question was left to the States, which independently created positive discrimination for OBC. The first States to implement reservations on the same basis as those for SC/ST were the four southern

[29]The main opposition was between Gandhi and Ambedkar, who had conflicting ideas about how to improve Untouchables status.

[30]This estimation comes from the Mandal Commission, which was established in 1979 with the mission of making a report on the status of the OBC. However, this number is controversial, because it has been calculated with figures from the 1931 census, the last census with details by jatis.

[31]The word "Central" is used to refer to what is defined at the federal level, and the word "State" to what is defined at the State level.

States (Kerala, Karnataka, Andhra Pradesh and Tamil Nadu) (Galanter, 1978) and the last were the States of the Hindu Belt (Rajasthan, Haryana, Uttar Pradesh, Madhya Pradesh). The Central Government finally caught up the movement by establishing in 1993 reservations for OBC in the Central administration and since 2008 27% of the seats in Central Universities[32] are reserved for OBC.

Therefore, there is now 49.5 % of the seats in higher education institutions reserved for low castes in India. This goes along with scholarships, free meals, etc. at lower levels of education. Whether it has been effective so far is however a matter of debate. For example Cassan (2011), who exploits the States borders' changes to evaluate the impact of affirmative action on the education level of scheduled castes, finds no impact. On the contrary, Bertrand et al. (2010) who look at the economic status of low castes who got into higher educational institutions thanks to quotas find that low castes who benefited from affirmative action are richer than those who did not.

2 How do others matter?

In the previous section, I have described how human capital in India is closely related to the institution in which social capital is created, the caste, and what are the mechanisms which are designed to break this historical situation where high castes have a lot of human capital and low castes have very little. In this section, I want to further explore the relationship between social capital and human capital, by broadening 1) the geographical focus, and 2) the question. More precisely, this section goes back to the question asked at the beginning of this introduction: *how do others matter in relation to human capital?* The literature in relation to this question is explored looking at two different perspectives, which are those that I consider in this dissertation: how does the human capital of others influence productivity? and how do others have an impact on human capital formation?

2.1 How does the human capital of others influence productivity?

The human capital of others has an impact on individual or aggregate productivity through three channels. The first channel relates to knowledge spillovers. Individuals learn from their co-workers or neighbors who have more human capital, which in turns increases their individual productivity. The second and third channels are

[32]In India, there are two kind of universities: the "Central Universities", which depend on the Central government of India, and the "State Universities" which are administered by State governments.

composition effects of human capital. The second channel depends on the complementarity between different levels of human capital (Kremer, 1993). If tasks and different levels of education are complementary in the production function, then the way the education is distributed in firms or in an economy as a whole has an impact on aggregate productivity. The same is true for the third channel. If returns to human capital are not linear in the level of human capital (Elbers and Gunning, 2004), which means that the marginal returns of human capital is not the same for low and high levels of education, then again, the way the education is distributed in an economy has an impact on aggregate productivity.

The first channel has therefore an impact on individual and aggregate productivity, while the impact of the second and third channels is only visible at the aggregate level. The empirical estimation of the impact of the human capital of others in the economy therefore follows two paths: one kind of literature estimates the impact of the composition of human capital in the economy on growth or productivity at the aggregate level, in which case there is no distinction between the three channels. The second kind of literature looks directly at the impact of co-workers or neighbors on individual productivity and is therefore only interested in estimating the impact of the first channel, knowledge spillovers.

The literature at the aggregate level focuses on the impact of *inequality* of education on productivity and growth. It is a fairly recent literature, which finds contradictory results. Londono and Birdsall (1997); Lopez et al. (1998), Castelló and Domenech (2002) and Castelló-Climent (2010b) find a negative impact of education inequality on income growth or income per capita. On the contrary, the impact is positive for Schipper and Hoogeveen (2005), Park (2006) and Rodríguez-Pose and Tselios (2009). The reasons for this discrepancy in the results will be explored in chapter 1. However, what can already be said from this list of results is that further research on this issue is required in order to understand what is the impact of the distribution of education on the economy of a country.

The literature specifically considering education spillovers is much more developed.[33] Here I only focus on the papers that I consider as the most important, because of their identification strategy or the context in which they study education spillovers.

As mentioned before, education spillovers can be observed at the aggregate level. This is what Acemoglu and Angrist (1999), Ciccone and Peri (2006) and Moretti

[33] A good overview of the research on this issue, as well as on the impact of the surrounding human capital on other issues is provided in Moretti (2004b).

(2004a) do, by comparing wages of workers in cities with more or less human capital. Acemoglu and Angrist (1999) do not find any impact, but Ciccone and Peri (2006) and Moretti (2004a) find that a higher share of college graduates in a city increases workers productivity.

A positive result is also found when education spillovers are studied at the plant level. Martins and Jin (2010), who use panel data on firms in Portugal, and Moretti (2004c), who looks at firms in the US, find important knowledge spillovers. The evidence is not as clear in the literature looking at education spillovers in a non-industrial context. Three papers study the impact of neighbors' human capital on agricultural productivity. Appleton and Balihuta (1996) and Weir and Knight (2007) find that the surrounding human capital matters for productivity. On the contrary, Asadullah and Rahman (2009) find no impact of neighbors' education on agricultural productivity. Again, these results, and the reasons why there is no consensus will be discussed in more details in chapter 2.

Among the papers previously quoted, none on them specifically focuses on India. Most papers at the aggregate level are concerned with the United States or developed countries in general. The exception is Lopez et al. (1998) who use panel data on 12 developing countries between 1970 and 1994. At the microeconomic level, the literature looking at education spillovers in plants also only uses data from developed countries. However, the literature studying agricultural productivity focuses on developing countries: Appleton and Balihuta (1996) and Weir and Knight (2007) look at education spillovers in farm productivity in Africa, and the geographical focus of Asadullah and Rahman (2009) is Bangladesh. The only related literature which is specifically about India is the one studying peer effects in the adoption of new technologies in agriculture. India is indeed a country where the Green Revolution rapidly expanded and succeeded. Worth mentioning are the papers by Foster and Rosenzweig (1995), who find important learning effects in the adoption of high yielding grains in India during the Green Revolution, and by Munshi (2004) who finds that learning from neighbors is weaker when the productivity of the new technology depends on unobservable individual characteristics.

The fact that others influence productivity, at the individual or aggregate level, is one aspect of the interaction between human capital and social capital. However, maybe even more important is how social capital interacts with human capital *ex-ante*. In other words, how does social capital influence human capital formation? The next section reviews the literature dealing with this question.

2.2 How do others influence human capital formation?

The question of how others influence human capital formation was actually the first question asked when the concept of social capital was created. The seminal paper on social capital of the sociologist Coleman (1988) is indeed entitled "Social capital in the creation of human capital". This paper is interested in how social capital in the family and in the community influences the creation of children's human capital and underlines important correlations between measures of social capital and dropout rates. But more importantly, this paper introduces the concept of social capital and opens the way to further research on this issue. Since this first paper, research on the topic has exploded, although most of the literature only considers social capital in a very reduced meaning: the core of the literature concentrates on the impact of peers' educational performance on children's performance, where performance is measured with test scores or dropout. Sacerdote (2011), who provides a review of this literature, underlines that there is a consensus on the existence of peer effects in educational outcomes.

Maybe more relevant to the topic on hand, is the literature which studies neighborhood effects on children's school achievement. In this literature there is no consensus on the fact that neighbors' characteristics have an impact on children's performance. For example, Goux and Maurin (2007) find that the proportion of uneducated families in the neighbors has a positive impact on the probability of repeating a year. On the contrary, Gibbons et al. (2013) find no impact of changes in the neighborhood composition on test scores. According to Patacchini and Zenou (2011), this variety of results can be explained by the fact that the impact of the neighborhood is complex, and may not be linear. This is indeed what they find when they study the impact of the quality of the neighborhood on children's education. They find that neighborhood quality and parents' involvement are complementary and neighborhood quality matters more for lower educated parents.

Taking a broader perspective, peers and neighbors seem to also have an impact on education related choices. De Giorgi et al. (2009), studying the choice of major of students at the University of Bocconi, find that they are highly influenced by the choice of their peers, leading to a mismatch between skills and qualifications in the labor market. Similarly, Marmaros and Sacerdote (2002) find that the probability that students take high paying jobs increases with the mean parental income of their dormmates during their first years as freshman. However, they are not able to differentiate if this is because students were influenced by their peers, or if they get concrete help from their peers to get a high paid job. But again, there seems to be no consensus on the question: Arcidiacono and Nicholson (2005) find that there are little peer effects in the choice of speciality for American medical school students.

There is also an extensive literature on how others influence opportunities, through their impact on employment outcomes. As underlined earlier, opportunities have an indirect effect on human capital formation, because people invest in education according to their expected returns. It has been shown that people notably use their network to get information about jobs (for empirical evidence, see for example Corcoran et al. (1980), Topa (2001) or Wahba and Zenou (2005)). Social networks are also useful because they play the role of referal (Topa, 2011). Consequently, networks decrease unemployment duration (Cingano and Rosolia, 2012).

How does this literature relate to India? Again, none of the papers previously mentioned is specifically concerned with India, and very few are about developing countries. However, what is interesting to see is that a growing literature considers that the ethnic group is where social capital is created. Little is still known about the role of ethnic networks in the direct creation of human capital, because ethnicity is usually taken into account by including dummies. But the literature begins to understand the role played by ethnic networks in jobs and carrier opportunities (see for example Munshi, 2003; Edin et al., 2003; Patacchini and Zenou, 2012). One interesting paper which directly makes the link between job opportunities and human capital investment is the one by Munshi and Rosenzweig (2006), mentioned above. They show that ethnic networks may not have always a positive impact. In particular here, they find that caste networks are possibly inefficient because there was little reaction in terms of educational investments for low castes to an increase in returns to education.

3 Outline of the thesis

This dissertation fits into the literature just mentioned by studying how productivity at the aggregate level and at the individual level is impacted by the human capital of other people in the economy, and how human capital creation is driven by social capital, directly or indirectly through available opportunities. The structure broadly follows the outline chosen to expose the literature. **Chapter 1 and 2** look at the link between productivity and others' human capital, and **Chapter 3 and 4** study the impact of caste on application to affirmative programs, which make the access to universities (chapter 3) and to jobs in the public sector (chapter 4) easier for low castes. Although the order of the four chapters can appear counterintuitive at first sight (one could expect to begin with the *creation* of human capital and to end with the *impact* of human capital), this organizational choice is actually driven

by how "others" are defined in the four chapters. The direction of this dissertation goes towards a narrowing of the definition of "others". In chapter 1, "others" is defined broadly and refers to all other workers in the economy. In the subsequent chapters, "others" refers to individuals' group, where the group under study is the individual's *varna* in chapter 2 and the individual's *jati* in chapter 3 and 4. Similarly, the perspective is narrowing: chapter one is a macroeconomic work on Indian States, whereas chapter 2 to 4 have a microeconomic focus.

The objective of **chapter 1** is to provide a macroeconomic perspective on how the human capital of others matter for aggregate productivity. As underlined in the literature review, when considering this issue researchers have focused on the impact of the inequality of human capital. This is also the approach that is taken here. More precisely, this chapter empirically analyses the relation between the distribution of education and income per capita of Indian States with six rounds of households' surveys, the NSS data, between 1983 and 2009-2010. Using two different measures of the distribution of education, a Gini of education and a Theil index, and dealing with the high correlation between the mean education level and its distribution by separating the "level effect" from the "composition effect" in the Gini coefficient, this chapter provides evidence that there is a negative relation between the *equality of education* and income per capita. This relation is however not linear and depends on the level of development: the relation is negative for richer States, but poorer States have an income per capita positively correlated (but not always significant) to the equality of education. As previously mentioned in section 2.1, the channels which may explain this impact are threefold: there may be education spillovers, the returns to education may be non-linear, and/or there may be complementarity between workers education level. In a second part, I consider these three channels successively. Data are limited - they do not cover the whole period- so too much emphasis should not be put on the results. But I find that all the three channels seem to play a role in explaining the relation.

Chapter 2 is a direct extension of chapter 1. It also explores how the human capital of others matter for productivity, but this question is now studied taking a microeconomic perspective. The focus is also more restrictive: in this chapter, the area of study is *rural* India, the reason being that the productivity which is taken into consideration is *agricultural* productivity. The objective is to analyze if there are education spillovers in rural India, by evaluating the impact of neighbors' education on farm productivity. To overcome the identification problem related to the endogeneity of network formation, I exploit the fact that social interactions mainly occur along caste lines in rural India. Defining neighbors as members from

the same caste also permits to rule out that the measured impact of members from the same caste is a spurious correlation due to unobservable characteristics.[34]

Using a household survey representative of rural India, The ARIS-REDS data, the results show that education spillovers do exist: one additional year in the mean level of education of members from the same caste increases households' farm productivity by 5%. The impact of neighbors' education decreases with households' level of education, showing that neighbors and households' head education are substitutes. I also find that education spillovers vary across crops. They are lower for rice than for wheat. This may be due to the fact that learning from others in rice is more complicated, because rice productivity depends a lot on fields' characteristics.

After having shown that the human capital of others has an impact on productivity, the rest of the dissertation is concerned with the broad question of how others matter in the creation of human capital. In **chapter 3**, I study the determinants of households application for affirmative action programs in education in India, with a focus on the role of social stigma, where social stigma is defined as the disutility arising from participation in a program (Moffitt, 1983). Affirmative action enables low castes to have an easier access to higher education through quotas in public institutions. To measure stigma, I look at the Other Backward Classes (OBC) and I analyze the impact of the status of households' subcaste on their probability of applying for reservations. If there is stigma, we expect households from castes with a higher social status to have a lower probability of applying for reservations. The identification strategy is based on the fact that the OBC group is composed of subcastes which are very different in terms of social status. The status of a subcaste group in rural India is locally determined and is strongly related to the proportion of land owned by this subcaste in the village. I use this exogenous and historical variation of status to identify the stigma effect.

I find that the probability of applying decreases with the social status of the caste group from which the household is from. This result can be interpreted as a stigma effect which leads households from locally high ranked groups to apply less for reservations in higher education.

Finally, **chapter 4** is concerned with affirmative action programs in the public sector, which provide reserved jobs for low castes in the administration. These programs, by changing the access of low castes to jobs that were in the past preempted by high castes, also changed low castes expected returns to human capital. However, the goal here is not to directly assess if those quotas had the effect of increasing low

[34]This is true under certain hypotheses which are discussed in more details in the chapter. This strategy is the one used in Munshi and Myaux (2006).

castes human capital. I consider reserved jobs under a slightly different, but related, angle: I am interested in the *concrete access* to those reserved jobs. Indeed, for the actual expected returns to education to be higher, low castes need to be concretely able to get the jobs. However, while any low caste can apply, the recruitment is highly discretionary and the recruiters can hire who they want (Chandra, 2004). Therefore, intermediaries and networks play an important role in providing access to those jobs. In this chapter I am looking at the impact of caste networks in getting access to reservations in the public sector. More precisely, I exploit political reservations for low castes to study the impact of having an elected local leader on low caste households' applications for reserved jobs in the public sector. Using data from 37 villages from three South Indian States from the ARIS-REDS survey, I find that households apply significantly more for reserved jobs when the council president is from their caste group. This seems to be due to a "patronage effect", where the council president uses his connections to help his caste-fellows to get a reserved job. These results show that access to reserved jobs is uneven across caste groups, because members from well connected castes have a facilitated access to higher levels of the administration. The expected returns and consequently investments in human capital are therefore likely to be different across caste groups, depending on their characteristics.

Overall, this thesis contributes to the growing literature which moves the focus from the private individual to the group, by looking at the interactions happening between human capital and castes. It provides some empirical support to the idea that human capital is related in several dimensions to social capital in general and to castes in particular. It therefore strengthens the idea that caste is still an unavoidable institution in India. Additionally it sheds some light into two important questions which are currently heavily debated in India. Chapter 1 and 2 contribute to the debates on the growing role of private education in India. These chapters, by showing that there are external returns to education, emphasize the need for a coherent education policy which does not only rely on private schools. Chapter 3 and 4, which find that application to quotas for low castes in higher education and in jobs in the public sector depends on several caste characteristics and notably on caste power, call for a large scale evaluation of the efficiency of affirmative action programs based on caste membership criteria.

Chapter 1

Distribution of education and income per capita: empirical evidence from Indian States

1 Introduction

In 2009-2010, the working-age population of Karnataka and Assam, two Indian States, had on average the same level of education (respectively 6.2 and 6.1 years of education). However, the pattern of distribution of their human capital was very different. Whereas in Karnataka 31% of the population was illiterate but 10% had a tertiary education, in Assam, only 18% was illiterate and 4% had a tertiary education. What are the consequences of different distributions of human capital? Which education policy favors economic development? The necessity of increasing human capital is not anymore a matter of debate, but the way it should be done is still an open research question.

Of course, the answer to what should be done depends on what outcome we have in mind. Development can mean higher growth and higher income per capita for some, and equality of opportunities for others. The answer will strongly differ depending on what one is thinking about. For example Castelló-Climent (2008) finds that a more equal distribution of human capital is good for democracy, whereas Park (2006) finds that a greater inequality of education favors income growth. The goal of this paper is not to look at the impact of the distribution of education on all the facets of development. In this paper, I only discuss the relation between distribution of human capital and income per capita, using data from Indian States between 1983 and 2009-2010.

The direction of the relation between distribution of education and income per capita is *a priori* not clear. The theoretical literature underlines three channels through which the distribution of education is related to income per capita and the three channels have contradictory impacts.

The first channel is that the returns to education may not be linear in the level of education (Elbers and Gunning, 2004). If this is the case, the impact of the distribution of education on income per capita depends on the returns to education in terms of productivity. If the returns' curve is concave in education, the marginal productivity of workers with high level of education is lower than the marginal productivity of workers with low level of education. Therefore, a more equal repartition of education has a positive impact on income. On the contrary, if the returns' curve is convex, inequality of education has a positive impact on income. Recent empirical evidence has underlined that the curve of returns to education in terms of wages is convex in India (Colclough et al., 2010).[1] If wages are equal to the marginal produc-

[1]Due to data availability, these estimates are only based on wages estimations. Since a large part of individuals are self-employed, the return to education could be different when taking into account this population.

tivity of workers, then in India a more equal distribution of education is expected to have a negative impact on income per capita.

The distribution of education also matters if there are externalities of education, through learning spillovers. Learning spillovers are usually assumed to go from high educated workers to low educated workers (Moretti, 2004b), in which case a less equally distributed human capital should have a positive impact on income per capita.

Finally, if workers with different education levels are complementary in the production process, then the distribution of education also matters. Kremer (1993) underlines that when tasks are also complementary, the optimal matching of workers is when workers of similar skills work together. This matching is achieved in labor markets without frictions. However, in labor markets which are not perfect, the distribution of education matters for income per capita, because a more unequal distribution of human capital lowers the probability that workers of similar skills can be matched together. Therefore, according to this channel, a more equal distribution of education should increase income per capita.

To summarize, the two first theoretical channels predict a negative relationship between *equality of education* and income per capita in India, whereas the last channel predicts a positive relationship. The overall relation between education distribution and income per capita is therefore an empirical question.

However, empirically there is still no consensus on the direction of the relation. Londono and Birdsall (1997), Lopez et al. (1998), Castelló and Domenech (2002) and Castelló-Climent (2010b) find a negative correlation between *education inequality* and income growth or income per capita. On the contrary, Schipper and Hoogeveen (2005), Park (2006) and Rodríguez-Pose and Tselios (2009) find a positive relation.

It is hard to point out what could drive this variety of results, because the comparison is difficult due to the diversity in the geographical coverage of those studies. For example, Lopez et al. (1998) focus on a panel of developing countries, Castelló and Domenech (2002) mix developing and developed countries and Rodríguez-Pose and Tselios (2009) only have developed countries in their sample.

Several hypotheses can still be formulated. The absence of consensus could be explained by methodological issues. Except for Schipper and Hoogeveen (2005), who study the impact of education inequality in rural Uganda, the identification of the impact in this literature only relies on cross-sections or large panels of countries. The heterogeneity that exists between these different countries is obviously hard to be taken into account and could drive the results that the authors get.

This literature also ignores that the distribution of education is highly correlated

to the mean level of education (Thomas et al., 2001).[2] Multicollinearity can make it difficult to uncover the partial effect of each variable (Wooldridge, 2003), and can also generate spurious estimates when the two highly correlated variables have a small explanatory power (Chatelain and Ralf, 2012).

Finally, the same pattern of results could be obtained if the relation between human capital distribution and income per capita is not linear and depends on the level of development as underlined by Castelló-Climent (2010b). If this is the case, the effect may vary a lot depending on the chosen sample of countries.

This paper therefore aims at bringing complementary evidence on the relationship between education distribution and income per capita, while improving previous literature on a methodological point of view. To avoid the great heterogeneity that exists across countries, I focus on a single country, India and use panel data on Indian States over 25 years. I deal with the high correlation between the mean education level and its distribution by using an equality index of education built following a methodology proposed by Berthélemy (2006). I find that there is a negative relation between the equality of human capital and the income per capita of Indian States. The relation differs depending on the level of development. The relation is negative in richer states, but positive (although not always significant) in poorer States. This result gives credit to the hypothesis that non-linearities may drive the diversity of results found in the literature.

The second objective of this paper is to explore the channels which drive the relation. While the literature has explored some channels that lead to an impact of education distribution on *growth* (Castelló-Climent, 2010a), there is no paper looking at the relation between education inequality and aggregate income or *income per capita*. Given the small amount of data available, one has to be careful on the conclusions. Nonetheless, the three channels previously underlined seem to play an important role in explaining the relation between distribution of education and income per capita.

The methodological improvements and the non-linearity results are the main contribution of this paper to the literature. However, this preliminary exploration of the channels is also informative, and nothing in that direction had been made so far.

The rest of the chapter is organized as follows. Section 2 explains the distribution variables used and in particular the construction of the education equality index. Section 3 and 4 present the data used and the empirical methodology. The results are shown in section 5 and section 6 empirically explores the theoretical channels.

[2]This point will be further explained in section 2.

Section 7 concludes.

2 Measures of the distribution of education

In the literature on the distribution of education, two different kinds of measures are used. The first category includes sample moments of the distribution, such as the variance of schooling. The advantage of the variance is that it is simple to compute. However, it does not have the properties that are required from a good measure of inequality. In particular, it only measures the inequality in absolute terms (it is not bounded) (Thomas et al., 2001) and it is not scale invariant. The second type copes with these limitations by notably measuring the inequality in relative terms. The most widely used are the Gini of human capital and the Theil Index. I focus on these two indicators in this paper.

2.1 The Gini coefficient of human capital

The Gini coefficient of human capital can be calculated as follows (Castelló and Domenech, 2002):

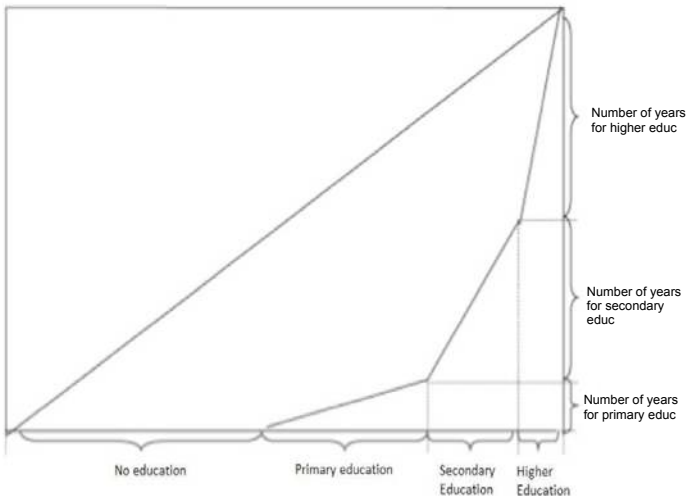
$$G_H = \frac{1}{2\overline{H}} \sum_{i=0}^{n} \sum_{j=0}^{n} |\hat{x_i} - \hat{x_j}| \, n_i n_j \tag{1.1}$$

Where G_H is the Gini of human capital, $\overline{H}$ is the mean level of education, i and j are the levels of education, $\hat{x_i}$ and $\hat{x_j}$ are the cumulated years of education for levels i and j of education and n_i and n_j are the share of people with a given level of education.

Like the classical Gini used to estimate the inequality of income, the Gini of human capital can be represented graphically with a Lorenz curve as in figure 1.1. The different levels of education are on the vertical axis and the cumulative proportion of people with a given level of education is on the horizontal axis. The Gini of human capital is two times the area between the Lorenz curve and the diagonal which represents perfect equality.

One issue with this index, as documented by Thomas et al. (2001), is that it is highly correlated with the mean level of education, and this correlation is higher in developing countries. The reason is twofold. First, the distribution of human capital is bounded: one cannot have more than a PhD level, that is to say more than approximately 20 years of education in most countries. Consequently, when the mean level of education increases, it cannot be the level of the most educated people which increases, but the level of less educated people, so the Gini of education

Figure 1.1: Lorenz curve for human capital with 4 education levels



Source : Berthélemy (2006)

automatically decreases. Second, the Lorenz curve is truncated along the horizontal axis (Thomas et al., 2001). As there is a large proportion of individuals in developing countries, a low mean level of education is associated with a high value of the Gini coefficient (see figure 1.1).

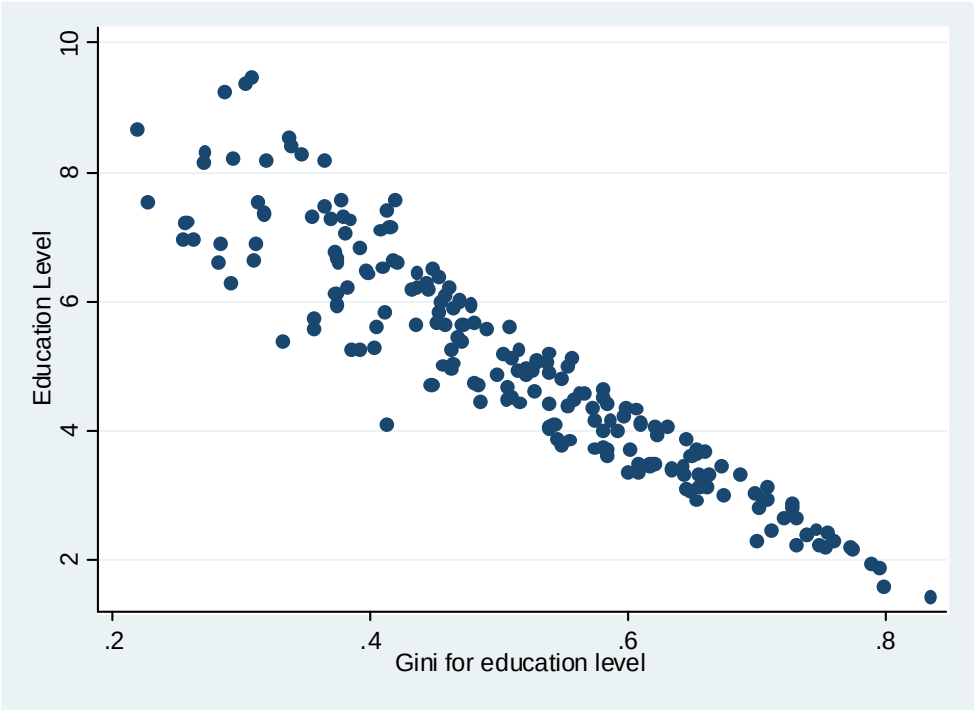
Figure 1.2 shows the relation between the Gini of human capital and the mean level of education in Indian States with six rounds of data between 1983 and 2009-2010. There is a strong negative correlation between these two variables, which seems to be even stronger for low levels of education.

2.2 Decomposing the Gini coefficient into its level and composition effect

This very high correlation between the mean level of education and its distribution complicates the estimation of the impact of the distribution of human capital. First, it creates econometric issues. Multicollinearity often provides too large standard errors (Wooldridge, 2003), or generate spurious estimates when the two highly correlated variables have a small explanatory power (Chatelain and Ralf, 2012). Second, it makes it difficult to understand if the measured impact of the distribution of human capital is actually a mean effect or a real distribution effect.

One solution is to disentangle in the Gini coefficient the level effect of education from the concentration effect. For this, I follow the theoretical work of Berthélemy (2006). Within a simple framework where there are only four different levels of

Figure 1.2: Education and Gini of human capital



education (no education, primary education, secondary education, higher education) and where each level has the same number of years, Berthélemy (2006) shows that the Gini of education can be written as a function of the mean level of education and a function Γ, where Γ is a measure of the *equality of a schooling system* and depends on its structure. Following his notations:

$$G_H = 1 - H\Gamma(\theta, \tau) \tag{1.2}$$

With

$$\Gamma(\theta, \tau) = 1 - 2\left(\frac{\theta + \tau + \theta\tau}{(1 + \theta + \tau)^2}\right) \tag{1.3}$$

G_H is the Gini of human capital, H is the aggregate level of human capital, θ and τ are respectively the ratio between the rate of secondary schooling and primary schooling and the ratio between the rate of tertiary schooling and primary schooling. As Γ depends non-linearly on θ and τ, it is hard to predict how it evolves with education policies. However, upon taking partial derivatives, several properties can be underlined. For not too large values of θ and τ, Γ decreases with θ and Γ always decreases with τ. Γ increases with θ for large values of θ.[3]

As shown in table 1.1, in the sample of States in the period under consideration, the average value for θ is 0.62. It means that on average the proportion of individuals

[3] Γ increases with θ when $\theta > 2(\sqrt{2}-1)$ and $(\theta - \sqrt{\theta^2 + 4\theta - 4})/2 < \tau < (\theta + \sqrt{\theta^2 + 4\theta - 4})/2$.

Table 1.1: Descriptive statistics for θ and τ

	Mean	Sd Dev	Min	Max
θ	0.62	0.34	0.22	2.2
τ	0.21	0.23	0.04	1.85

The values for θ and τ are computed over the sample of 162 observations.

with a secondary education is a little more than half the proportion of individuals with a primary education. τ is much smaller, with a mean value of 0.21. On average, the proportion of individuals with a higher education is one-fifth the proportion of individuals with a primary education.

How does the equality index varies with θ and τ? The values of θ and τ for all States at every period are provided in table A1.1 in the appendix. As we can see, most of the States have small θ. Therefore, the education equality index decreases with θ and τ in the majority of States. In few States θ is large and meets the criteria formulated in footnote 3 so the equality index increases with θ.

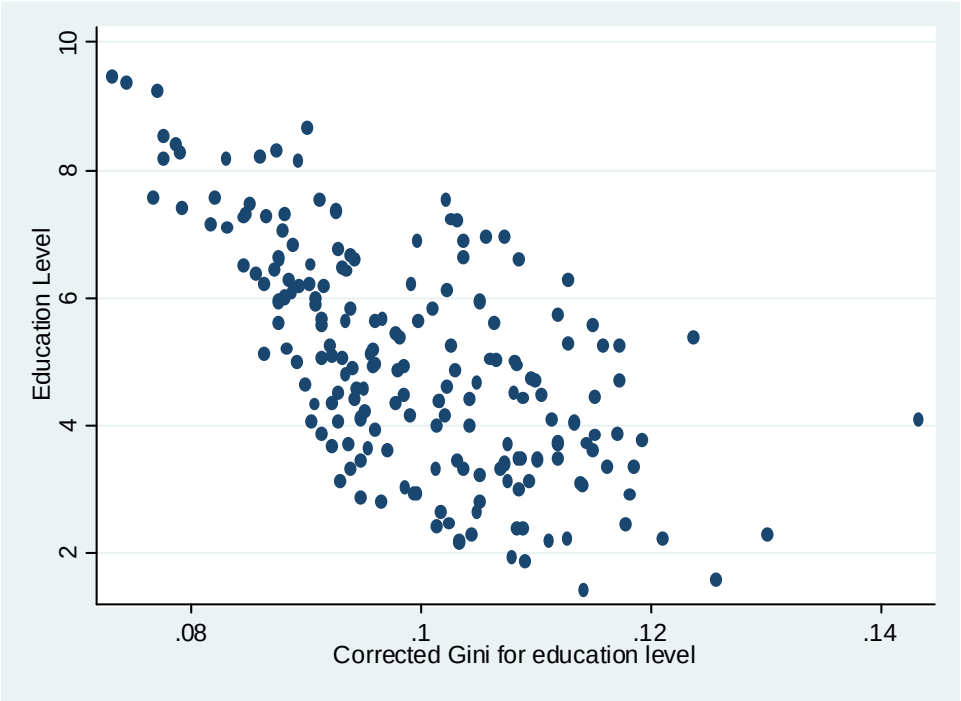
Empirically, Γ can be computed following equation (2): $ln\Gamma = ln(1-G_H) - lnH$.[4] Figure 1.3 shows the relation between the "education equality index" (Γ) and the mean level of education in my data. As we can see, the correction is not perfect: there is still some correlation between the mean level of education and the corrected Gini. One explanation to this persistent correlation, suggested by Berthélemy (2006), is that the ratios of relative schooling θ and τ, on which the education equality index depends, are related to aggregate human capital. The supply of education depends on the presence of skilled teachers in an economy, which itself depends on the aggregate human capital. But primary education depends less on the presence of skilled teacher than secondary education, which itself depends less on skilled teachers than tertiary education. This would explain why θ and τ, and therefore the equality index of human capital, is correlated to the aggregate human capital. However, the correlation is much lower and this specification of the Gini coefficient reduces multicollinearity issues in the estimation.

2.3 The Theil index of Human Capital

To make sure that the results are not driven by the use of the education equality index, I also use another inequality measure, the Theil index. The Theil index is

[4]To get a more precise definition of the mean education and of the equality indicator as well as to minimize measurement errors in the empirical specification, I use eight levels of education to compute the Gini coefficient of human capital. The different education levels are described in table A1.2 in the appendix. So this specification does not follow exactly Berthélemy (2006)'s theoretical model.

Figure 1.3: Education and Corrected Gini of human capital



widely used and has the advantage of being more sensitive to changes at the top of the distribution.[5] The formula of the Theil index is:

$$T_H = \frac{1}{N}\sum_{i=1}^{N}\frac{H_i}{\overline{H}}ln(\frac{H_i}{\overline{H}}) \tag{1.4}$$

where $\overline{H}$ is the mean education, H_i is the education level of individual i and N is the number of individuals in a given State.

3 Empirical specification

To estimate the relation between the distribution of education and income per capita I use the following specification:

$$lnY_{it} = \beta_1 ln\overline{H}_{it} + \beta_2 X_{it} + \beta_3 DISTR_{\overline{H}_{it}} + \beta_4 GINIcons_{it} + \gamma_t + a_i + u_{it} \tag{1.5}$$

where Y_{it} is the income per capita of State i at time t, $\overline{H_{it}}$ is the mean level of education of the State working-age population, and X_{it} is a set of control variables at the State level.[6] $DISTR_{H_{it}}$ is the education equality index or the Theil index

[5] It is therefore less correlated to the mean than the Gini, because it puts more weights to changes really related to the distribution.

[6] The proportion of workers in agriculture and in the manufacturing sector, electricity con-

depending on the specification.

As the distribution of human capital is related to the distribution of incomes, which may have an independent impact on income per capita, I also control for the income distribution in the State, which is proxied by a consumption Gini index, $GINIcons_{it}$ as it is commonly done for countries where income data are not very reliable. I control for shocks common to all States by adding time dummies represented by γ_t.

There are concerns about endogeneity. The fact that the variable of interest, the distribution of education is calculated on more than 15 years old individuals rules out any reverse causality problem. But the distribution of education may be correlated to factors also influencing income per capita. I control for time-invariant unobservables by adding State dummies a_i and standard errors are clustered at the State level. Robustness checks are also conducted with an alternative panel data method, the Blundell and Bond (1998) system GMM. To deal with endogeneity issues concerning the other control variables, all right hand side variables are lagged nine months[7]. Identification is further questioned in section 5.2.

4 Data and variables definition

4.1 Data sources

The data used to estimate the relation between the distribution of education and income per capita are from various sources. For the dependent variable, income per capita, I use data from the Reserve Bank of India. It provides the net State domestic products per capita. For several States, the data were not available for the whole period (1984 to 2010-2011).[8] Furthermore, three States have been split during the period.[9] So out of the current 35 States and Union Territories the relation between education equality and income per capita can only be estimated for 29 States and 162 observations.

The education and consumption variables, and the proportion of workers in agriculture and in the manufacturing sector come from six rounds of household

sumption per capita and State expenditures per capita.

[7]The choice of nine months is due to data availability.

[8]There are no estimates at all for Dadra and Nagar Haveli, Daman and Diu and Lakhshadweep; for Mizoram, there are no consistent estimates for 1984, 1988-89 and 1994-95; for Sikkim, Goa, Nagaland and Chandigarh for 1984 and 1988-89 and for Arunachal Pradesh there is no estimates for 1984. The States with at least three observations are kept because it is the minimum required to do a System GMM.

[9]Chhatisgarh has been created out of Madhya Pradesh, Jharkhand out of Bihar and Uttarkhand out of Uttar Pradesh. I treat these States as they were in 1983.

surveys aggregated at the State level. They are from the National Sample Survey Organization, that conducts households surveys approximately every five years. The National Sample Surveys (NSS) are very large surveys, conducted over more than 100,000 households and 600,000 individuals. Their size and the sampling strategy[10] make them representative at the below-State level. The rounds used here are from 1983, 1987-88, 1993-94, 1999-2000 and 2009-10.

Finally, data on State expenditure per capita are from the Reserve Bank of India, and electricity consumption per capita is from various reports of the Planning Commission of India.

4.2 Variables construction

The education equality index is calculated as explained in section 2.2. But to build it and calculate the mean level of education, I need to convert educational attainment into years of education. For this, I use the same classification as Kingdon and Theopold (2008). The conversion table is given in table A1.2 in the appendix. The mean level of education and the equality indicator are then computed using only the working-age population (more than 15 years old).

The household level consumption data are used to construct the consumption Gini. The formula is $GINIcons = \frac{2\sum_{i=1}^{N} iy_i}{N\sum_{i=1}^{N} y_i} - \frac{N+1}{N}$, where N is the number of households, and y_i is the consumption of household i.

Finally, the proportions of workers in agriculture and in manufacture are computed using the principal activity declared by individuals in the NSS data. The occupations are aggregated in three sectors, agriculture, manufacture and services.

5 Results

5.1 Estimation with State fixed effects

The results of the estimation in fixed effects with the education equality index are shown in table 1.2.

In the first column, I only include the mean level of education and the education inequality index. The coefficient on the education equality index is negative, which shows that the equality of education is negatively correlated to income, but the level

[10]The NSS use a stratified multi-stage design. States are first divided into agro-economic regions which are groups of contiguous districts, similar with respect to population density and crop pattern. The number of villages and town to be surveyed are then attributed to each region according to their population. In order to have adequate numbers of households from specific categories of the society, some categories are over-sampled, so the use of weights is required when aggregating the data. Further details on the sampling method are available at http://mail.mospi.gov.in.

Table 1.2: Variable of interest: Equality of education index estimated with FE

Dependent variable:	Net State domestic product per capita				
	(1)	(2)	(3)	(4)	(5)
Mean Educ (log)	-0.532**	-0.552**	-0.565**	-0.511**	-0.662**
	(0.240)	(0.231)	(0.224)	(0.233)	(0.254)
Education Equality	-1.029	-1.188*	-1.200**	-2.523***	-3.833***
Index (log)	(0.623)	(0.621)	(0.561)	(0.853)	(0.786)
Consump. Gini (log)		-0.207	-0.230	-0.286*	-0.338
		(0.143)	(0.158)	(0.153)	(0.222)
Prop. in Agr			0.361		
			(0.674)		
Prop. in Manuf			5.853		
			(3.831)		
Electricity cons.				0.130	
p.c. (log)				(0.109)	
State exp p. c.					-0.0319
(log)					(0.124)
Time dummies	Yes	Yes	Yes	Yes	Yes
$\beta_1 - \beta_2$	0.497	0.636	0.635	2.012	3.171
SE	0.566	0.571	0.534	0.955	0.960
N	162	162	162	90	99
r2	0.897	0.899	0.903	0.939	0.937

Robust standard errors clustered by State in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

of the coefficient is not very precisely estimated. The estimated coefficient on the mean education level is also negative, but one has to keep in mind that it is not the "true" coefficient, because lnH is also included in the education equality index. The true value of the coefficient is $\beta_1 - \beta_2$[11] and is reported at the bottom of table 1.2 along with its standard errors. It is positive but not significant.

In the second column, I add consumption inequality. It is negative but its level is not precisely estimated. However, its inclusion increases the absolute value of the coefficient of the education equality index, which turns significant at a 10% level, showing that the education equality index was actually capturing some income distribution effect.

In column (3), (4) and (5), other control variables at the State level are added. In column (3), I add the proportion of individuals working in agriculture and the proportion of people working in manufacture in order to take into account the industrial structure of the economy in the State. The reference group is the service sector. These additional variables are not significant and they neither change the level, nor the significance of the education equality index. It is also important to control for physical capital because it is a strong determinant of income per capita.

[11] It comes from the fact that $ln\Gamma = ln(1 - G_H) - lnH$.

As no capital data are available at the State level in India, I use electricity consumption per capita as a proxy in column (4). The data on electricity consumption were missing for several States and are not available for 1983, so the number of observations is dramatically falling in this estimation.[12] The same problem arises in column (5), with the addition of State expenditures which proxy for the amount spent on education that may be correlated with the structure of the education system and with income per capita. Given the reduction in the number of observations, the coefficients are not strictly comparable. But some results are still worth noting. First, the coefficient of the equality index is negative and significant, confirming the negative relation between education equality and income per capita. Second, the sign of the consumption Gini is still negative and more precisely estimated when the proxy for physical capital is added.

These results show that the equality of education is negatively related to income per capita over the period: a 10% increase in the equality of education index is related to a 13% decrease in income per capita. What does it mean concretely for the structure of schooling? As the the equality of education index is non-linearly related to θ and τ, which are respectively the proportion of individuals in the population with a secondary education over the proportion of individuals with a primary education and the proportion of individuals in the population with a higher education over the proportion of individuals with a primary education, the impact of a change in θ or τ depends on the values of θ and τ. If we take their mean value in the sample, a one point increase in θ increases income per capita by 26%. In other words, going from a structure of schooling where the proportion of individuals with a secondary education is a little more than half the proportion of individuals with a primary education to a situation where there are one and half more individuals with a secondary education than individuals with a primary education is related with an increase in income per capita of 26%. The impact of a change in τ is even bigger: an increases of one point in τ is related with an increase in income per capita of 98%. It means that going from a situation where there is approximately five time more individuals with a primary education than individuals with a higher education to a situation with a little more individuals with a higher education than with a primary education almost doubles income per capita. These numbers seem big, but in perspective with the investments required to achieve this kind of goal, they are actually not that impressive.

Table 1.3 shows the results when the distribution of education is measured with

[12] This is the reason why I do not use investment data or electricity consumption as a proxy in every estimation.

Table 1.3: Variable of interest: Theil index estimated with FE

Dependent variable:	Net State domestic product per capita				
	(1)	(2)	(3)	(4)	(5)
Mean Educ (log)	0.331	0.445	0.465	1.573	2.923**
	(0.537)	(0.546)	(0.512)	(0.911)	(1.037)
Theil index	0.849	0.980	1.021*	2.100**	3.627***
	(0.608)	(0.612)	(0.555)	(0.856)	(0.875)
Consump. Gini (log)		-0.190	-0.216	-0.249	-0.382*
		(0.142)	(0.159)	(0.161)	(0.217)
Prop. in Agr			0.352		
			(0.674)		
Prop. in Manuf			6.087		
			(3.738)		
Electricity cons. p.c. (log)				0.146	
				(0.111)	
State exp p.c. (log)					-0.0418
					(0.112)
Time dummies	Yes	Yes	Yes	Yes	Yes
Observations	162	162	162	90	99
r2	0.896	0.897	0.902	0.935	0.934

Robust standard errors clustered by State in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

a Theil index. The Theil index is an indicator of *inequality*, so we expect the coefficient on this variable to be positive. Table 1.3 confirms the previous results: a more unequal distribution of human capital is associated with a higher income per capita. As using the Theil index does not significantly change the results, in the next tables I only show the results with the education equality index.

5.2 Robustness Checks

Section 4 finds that income per capita and equality of education are negatively related in Indian States. Although I control for State unobservables by estimating the relation with fixed-effects, the point estimate may still be biased if there is selective migration. In the first subsection I discuss the bias induced by migration and in the next subsection I provide evidence that the results are not driven by sample or estimation choices.

5.2.1 Migration

One important endogeneity problem arises if migration affects the distribution of human capital non-randomly, for example if States with higher income per capita attract more educated migrants. In this case, the estimated negative relation between equality of education and income per capita is actually a spurious correlation

due to higher educated people migrating to richer States.

Migration rates in India are indeed higher than what is usually thought. According to the census data, the percentage of individuals residing in a place other than their place of birth has been quite stable between 1981 and 2001, around 30%. However, when only looking at inter-State migration, the figures draw a different picture: only 6.5% of the more than 15 years old individuals were born outside of their State of residence in 2001, as reported in table A1.3 in the appendix. Therefore, although inter-State migration is increasing,[13] the number of inter-State migrants is still very low and the bias induced by selective migration should be small. Additionally, only one-fourth of the migrants declare having migrated for work related reasons,[14] while the other migrations are driven by familial reasons (Marriage or moved with households). While people migrating for work may select the State where to migrate depending on its economic strength, there is no obvious reason why this should be the case for familial motivated migrations.

There are several explanations to this low migration rate. First, States boundaries broadly follow linguistic zones and linguistic barriers prevent individuals from moving from one State to the other. The caste system is also often presented as a constraint to migration. Munshi and Rosenzweig (2009) for example show that the insurance system that exists within castes prevents people from migrating far away from their place of origin.

Table 1.4: Migration robustness checks

Dependent variable:	Net State domestic product per capita			
	(1)	(2)	(3)	(4)
Mean Educ (log)	-0.636***	-0.805***	-0.508*	-0.424*
	(0.226)	(0.261)	(0.261)	(0.212)
Education Equality	-1.790**	-2.024***	-1.228*	-1.081*
Index (log)	(0.651)	(0.684)	(0.651)	(0.575)
Consump. Gini (log)	-0.224	-0.307	-0.163	-0.222
	(0.179)	(0.211)	(0.150)	(0.139)
Time dummies	Yes	Yes	Yes	Yes
$\beta_1 - \beta_2$	1.154	1.219	0.720	0.657
SE	0.685	0.781	0.623	0.507
N	140	121	152	162
r2	0.913	0.911	0.893	0.897

Robust standard errors clustered by State in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

However, even if selective migration is small, it may still have a non-marginal

[13]The data are not available for the more than 15 year old population, but the proportion of inter-State migrants in the population was 4.1% in 1991 and 4.7% in 2001.

[14]The census does not give the reasons for migration from the place of birth, so these estimates are for migration from last place of residence.

impact on the coefficient of the education equality index. So table 1.4 provides some robustness checks. In column (1) I exclude Andaman and Nicobar Islands, Delhi, Chandigarh and Pondicherry. These four "Union Territories" have a particular administrative organization and because of their small size they also have very important migration movements. In column (2) I exclude States where migrants from other States or countries in 2001 constitute more than 5% of the population. These States are the four Union Territories previously mentioned plus four other States (Arunachal Pradesh, Goa, Haryana and Sikkim). In column (3), I exclude States that have an out-migration rate over 5%, namely Delhi and Chandigarh. Finally, given that the main type of inter-States migration is from rural to urban areas, in column (4) I calculate the education variables (mean and distribution) with only the population living in the rural parts of the States. In every specification, the coefficient on the distribution of education is negative and strongly significant. Interestingly, the exclusion of States with high migration rates increases the size of the coefficient. Therefore, if any, the bias due to selective migration seems to lower and not increase the estimated impact of the education inequality index.

5.2.2 Additional robustness checks

Table 3.4 presents complementary robustness checks. Again, I just present the results with the education equality index. Column (1) uses an alternative panel data method to estimate the relation and columns (2) and (3) are estimated with changes in the sample and in the definition of the variables.

Column (1) presents the estimations with an alternative panel data method, the Blundell and Bond (1998) system GMM which takes into account the potential endogeneity of the education variables (mean level and distribution) and of the consumption Gini. The choice of System GMM is made over first-differenced GMM because System GMM performs better when the number of observations is small (Blundell and Bond, 1998). The methodology of System GMM is as follows: the unobserved heterogeneity between States is controlled for by first-differencing the equation. Then the differenced-equation is instrumented with internal instruments: the differenced endogeneous variables are instrumented by their lagged values. This differenced equation is simultaneously estimated with the equation in level where the variables are instrumented by their differenced lagged values. The use of this methodology requires three hypotheses to be valid. First, the residuals should be uncorrelated at the second order. Second, the lagged values of the control variables must be exogenous to their differenced value. Third, the addition of the equation in level, specific to the System GMM, requires that the differenced lag values of the control variables are uncorrelated with the fixed effects. To ensure the validity

Table 1.5: Robustness checks

Dependent variable:	Net State domestic product per capita		
	(1)	(2)	(3)
	System GMM	No NSS 55	4 levels
Mean Educ (log)	0.679**	-0.620**	-0.461**
	(0.260)	(0.245)	(0.190)
Education Equality	-2.453**	-1.237*	-0.889*
Index (log)	(0.957)	(0.687)	(0.453)
Consump. Gini (log)	0.480	-0.198	-0.169
	(0.361)	(0.172)	(0.135)
Time dummies	Yes	Yes	Yes
$\beta_1 - \beta_2$	3.132	0.617	0.429
SEround	0.886	0.628	0.456
AR(1) test	0.973		
AR(2) test	0.697		
Hansen J test	0.396		
Diff-in-Hansen test	0.557		
Observations	162	133	162
N. States	29	29	29
N. instruments	30		
Instr lags	t-3 to t-4		

Robust standard errors clustered by State in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

of these three hypotheses, various tests are used: the Hansen test of overidentifying restrictions controls for the exogeneity of the whole set of instruments. The difference-in-Hansen test controls for the exogeneity of the instruments used for the equation in level and the Arellano-Bond test indicates the autocorrelation of the residuals. All the variables are considered as endogenous. The results confirm the findings on the education equality index: it is negatively and significantly related to income per capita.

In column (2) the estimation is made without the data from 1999-2000. The questionnaire on consumption for this survey was changed and data are said to be not strictly comparable. Even if it only affects the Consumption Gini variable, as this variable is correlated with the equality index, it might change the results. Again, the results are not affected by this change.

Finally, in column (3) the education variables (mean and equality index) are recalculated with only four levels of education, in order to strictly follow Berthélemy (2006)'s theoretical work. Even if the coefficient on the education equality index is slightly smaller than before, the results broadly confirm the previous findings.

5.3 Non-linearities in the relation between education equality and income per capita?

The results above show a significant and negative association between the equality of education and income per capita. In this section, I check if the level and the direction of the relation depend on the period considered or on the level of development of the State. To preserve the size of the sample, these issues are tested using interaction terms.

I first test if the effect varies over time by creating a dummy variable which is equal to 0 if the education equality index is observed during the three first periods of the sample, that is to say in 1983, 1987-88 and 1993-94 and 1 if the variable is observed in the three most recent periods, in 1999-2000, 2004-05 and 2009-10. The intuition is that the effect of the distribution of education can be different before and after the liberalization that happened progressively from 1991 onwards, in particular because returns to education may have changed during this period. This dummy variable is then interacted with the education equality index. Column (1) and (2) in table 1.6 shows the results.

Table 1.6: Non-linearities estimated with FE

Dependent variable:	Net State domestic product per capita			
	(1)	(2)	(3)	(4)
Mean Educ (log)	-0.526**	-0.546**	-0.226	-0.244
	(0.250)	(0.241)	(0.242)	(0.232)
Education Equality	-0.924	-1.085*	0.875*	0.722
Index (log)	(0.627)	(0.628)	(0.464)	(0.470)
Educ Equality*After	-0.201	-0.194		
1999	(0.468)	(0.437)		
Educ Equality * Over			-1.663***	-1.687***
Median			(0.399)	(0.396)
Consump. Gini (log)		-0.206		-0.234*
		(0.143)		(0.115)
Time dummies	Yes	Yes	Yes	Yes
Joint significance	0.284	0.185	0.000628	0.000584
N	162	162	162	162
r2	0.898	0.899	0.915	0.917

Robust standard errors clustered by State in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The results are similar to previously. The coefficient on the equality of education index is negative and significant when the consumption inequality is controlled for. But the interaction term is not significantly different from zero: it consequently

seems that the relation between the education equality index and income per capita is the same during the first ten years than during the most recent ten years.

The second non-linearity test is on the level of development. Does the relation vary depending on the level of income per capita? I test this issue by creating a dummy variable which indicates if the State is over the median or under the median of income per capita in the sample. The dummy variable is equal to 0 if the Net Domestic Product of the State in 1999-2000 was inferior to the median Net State Domestic Product this year and 1 otherwise. The date of 1999-2000 is chosen because it is in the middle of the sample.[15] This dummy variable is then interacted with the education equality indicator.

Column (3) and (4) in table 1.6 shows the results. Here the results underline a strong heterogeneity in the relation. The coefficient on the interaction term is negative and significant at a 1% level and the coefficient of the education equality index even becomes positive (but only significant in the first specification). Moreover, the p-value for the joint significance of the education equality index and the interaction term (shown under the tables) strongly rejects the null hypothesis that these two coefficients are jointly zero. Therefore, it seems that in States with a low income per capita the education equality index is positively correlated with income per capita, whereas in States with a higher income per capita the relation is significantly negative. This result provides a key for reading the results from previous literature. Indeed, when looking at the composition of the samples, it seems that when developing countries are included in the sample, education *inequality* has a negative impact on growth or income per capita (Lopez et al., 1998; Castelló and Domenech, 2002), whereas when the sample is only composed of developed countries, education inequality has a positive impact (Rodríguez-Pose and Tselios, 2009).

6 Why is the equality of education negatively related to income per capita?

In this section, I explore the channels which may explain the relation of the equality of education with income per capita. In the introduction, I explained three theoretical channels underlined by the literature: returns to education, externalities and complementarities.

[15] However, the results hold if the status of the State (over or under the median) is defined at each period with the median of the year.

If we think about income in terms of the output of a production function, the three channels can be related to two different concerns on the production function: how workers with different levels of education contribute to the production and what is the form of this production function. If workers with different levels of education contribute differently to the production, it means that the marginal return to education of a worker depends on its level of education. In this case the return to education is said to be non-linear and the distribution of education has an impact on income per capita. The form of the production function also determines the impact of the distribution of education. First, workers with different levels of education may be complements or substitutes in the production function. Second, there may be externalities, in which case the productivity of a worker with a given level of education depends on the level of education of its coworkers.

In the rest of the section, the different channels are explained in more details and are tested empirically.

6.1 Complementarity of similar workers in the production function

The first channel that I consider is the channel underlined by Kremer (1993).[16] Kremer (1993), under certain hypotheses and notably that workers with the same amount of human capital are complements in the production function, shows that the situation which maximizes income is a situation where workers with the same education level are matched together in firms.[17] However, this perfect matching can only happen if labor markets are perfect and workers fully mobile. If there is no perfect mobility, then the distribution of education matters because it affects the probability of workers with a same education level to be matched together.

To test this mechanism, I take advantage of Aghion et al. (2008)'s database which classifies Indian States depending on their labor market regulations. The idea is quite simple: the probability that similar workers are matched together in equilibrium depends on the capacity of the employer to hire similar workers, which in turn depends on labor regulation laws.[18] Therefore, in States were labor market regulations are more pro-worker, I expect a positive relation between the equality of education and income per capita. Relating to my results, as the equality of education is negatively related to income per capita, I expect that the relation will be

[16] But it is the third channel described in the introduction.

[17] This point is not demonstrated by Kremer (1993) but he refers to Becker (1991) for the mathematical proof.

[18] Crudely, the idea is that if the employer realizes that his workers are not matching, he can easily try to find a better match if labor market regulations are more pro-employer.

less negative in States were labor market regulations are more pro-worker.

The legislation on labor markets in India is principally governed by the Industrial Disputes Act of 1947. However, as industry legislation in India is a competence which is shared between the central government and States government, the legislation has been differently amended by the different States over time. Aghion et al. (2008) in their database code every State amendment as neutral (0), pro-worker (+1) or pro-employer (-1). The scores are then added up on the whole period (1947-1997), in order to have an idea of the regulation direction for each year of the period.

I use this information on labor legislation to study if the relation between the distribution of education and income per capita differs depending on the type of labor regulation. For that, I interact the education equality index with the level of regulation at that year. The coefficient on the interaction term is then estimated as previously, except that the number of periods is reduced to 4 because Aghion et al. (2008)'s dataset stops in 1997.[19] The number of States is also reduced to 17, which leaves us with only 64 observations. Results are shown in table 1.7.

Table 1.7: Equality of education and labor regulation

Dependent variable:	Net State domestic product per capita			
	(1)	(2)	(3)	(4)
Mean Educ (log)	-0.565**	-0.622**	-0.610***	-0.599**
	(0.201)	(0.221)	(0.183)	(0.275)
Education Equality	-2.222**	-1.537	-1.258	-3.821***
Index (log)	(0.975)	(1.213)	(1.037)	(1.011)
Educ Equality *		0.240	0.594	0.0839
Labor Regulation		(0.301)	(0.433)	(0.396)
Consump. Gini (log)	-0.258	-0.127	-0.0814	-0.294
	(0.295)	(0.327)	(0.333)	(0.236)
Labor regulation		0.573		
		(0.688)		
Time dummies	Yes	Yes	Yes	Yes
Joint significance		0.0326	0.0254	0.000582
N	64	64	64	96
r2	0.832	0.840	0.848	0.937

Robust standard errors clustered by State in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Column (1) shows the main results estimated with the reduced sample available with Aghion et al. (2008)'s data on labor regulation. The effect is slightly bigger than when estimated with the full sample and is more precisely estimated with a p-value of the coefficient on the equality of education at 5%. Column (2) shows

[19] As data are not available for 1999, I proxy the data for 1999 by those from 1997

Table 1.8: Impact of education equality

Type of regulation	Impact size	p-value
Pro-employer	-1.851** (0.825)	0.040
Neutral	-1.258 (1.037)	0.244
Pro-worker	-0.664 (1.358)	0.632

* $p < 0.10$. Standard errors clustered at the State level are in parentheses. The coefficients are calculated with the specification of column 3 in table 1.7.

the results with the interaction term. The coefficient of the education equality index is negative and the one of the interaction term is positive. They are both not statistically significant, but they are jointly significant as shown by the p-value reported at the bottom of the table. It therefore seems that *ceteris paribus*, in the States which have more pro-workers regulation (Labor regulation index > 0), the relation of the equality of education with income per capita is less negative.

However, it is a little hard to understand what the coefficients concretely mean, because the labor regulation index takes a large range of value: it goes from - 3 to + 4. To get a clearer understanding of the relation between education equality and income per capita given a certain regulation level, I simplify the labor regulation index. As States are never moving from a kind of labor regulation to another during the period under study (a State which has a pro-worker regulation never goes to a pro-employer regulation for example), I classify the States into only 3 categories: States which have a pro-employer regulation get a score of -1, States which are neutral have 0 and States which are more pro-worker get +1. The index for a given State is therefore fixed over time. The results with this index are reported in column (3). The regression results are similar to the results obtained in column 2. The education equality index is negatively related to income per capita, and the interaction term is positive. Table 1.8 clarifies the relation, by showing the coefficient for the three types of labor regulation: when States have a pro-employer regulation, the education equality index is strongly negative and significantly different from zero. However, the coefficient decreases (is less negative) as we move to a less pro-employer regulation, and becomes not significantly different from zero.

Finally in column (4), I use the 6 periods, with the hypothesis that the kind of labor regulation (pro-worker, neutral or pro-employer) did not change in the most recent period. The interaction term is smaller but still positive.

These results are not to be too much emphasized. First, the validity of the test relies on the hypothesis that pro-employers legislations allow firms to have a better matching than pro-workers legislations. Second, the small number of observations prevents from rushing to more global conclusions. However, although it does not prove that there is a matching story under the results, it at least gives some credit

to this channel.

6.2 Externalities and non-linearity in the returns to education

I now explore the two remaining channels, the channel of externalities and the channel of returns to education. These two channels are hard to distinguish, because they have the same cause and the same consequences. Knowledge spillovers go from more educated workers to less educated workers. So given a mean education level, an increase in the share of people with a tertiary education increases income per capita while education equality decreases. The impact is the same if we consider the returns to education channel. As the return curve seems to be convex in India (Duraisamy, 2002), we expect that an increase in the share of people with a higher education increases income per capita while the equality index decreases. Therefore in both cases, the negative relation between the education equality index and income per capita captures the positive impact of an increase in the share of people with a tertiary education given a mean education level.

I consequently consider these two channels together by looking at the consequence of adding the proportion of people with a higher education (graduate and above) in the regression on the coefficient on the equality of education. Table 1.9 shows the results.

The first column is the baseline regression. The second column shows the results when the proportion of higher educated people is added. The coefficient on the proportion of people with a high education level is positive and big even though it is not precisely estimated. This positive sign is not surprising: other things being equal, the higher the proportion of highly educated people, the higher income per capita. But what is more interesting is that when this variable is added to the baseline model, the coefficient on the education equality index drops (in absolute value) and becomes not statistically different from zero. It shows that the equality index actually captures some externality or non-linearity effect. To check if this effect is really due to the proportion of highly educated people, in columns (3), (4) and (5) I replace this variable by the proportion of illiterate, primary educated and secondary educated. In those three regressions, the coefficient on the education equality index has the same level than in the first regression and is statistically different from zero. Consequently the proportion of people in the other levels of education does not impact the education equality index.

It consequently seems that some part of the effect of the equality of education is

Table 1.9: Equality of education and higher education

Dependent variable:	Net State domestic product per capita				
	(1) Baseline	(2) Higher educ	(3) Illiterate	(4) Primary	(5) Secondary
Mean Educ (log)	-0.552**	-0.512*	-0.718*	-0.557	-0.576**
	(0.231)	(0.250)	(0.407)	(0.365)	(0.231)
Education Equality Index (log)	-1.188*	-0.751	-1.179*	-1.199	-1.097
	(0.621)	(0.715)	(0.626)	(0.805)	(0.645)
Consump. Gini (log)	-0.207	-0.187	-0.182	-0.208	-0.195
	(0.143)	(0.137)	(0.150)	(0.141)	(0.138)
prop higher educ		1.613			
		(1.250)			
prop illiterate			-0.507		
			(0.792)		
prop primary educ				0.0177	
				(0.713)	
prop secondary educ					0.245
					(0.842)
Time dummies	Yes	Yes	Yes	Yes	Yes
$\beta_1 - \beta_2$	0.636	0.240	0.462	0.643	0.521
SE	0.571	0.611	0.613	0.606	0.605
N	162	162	162	162	162
r2	0.899	0.900	0.900	0.899	0.899

Robust standard errors clustered by State in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

due to the positive impact of the proportion of highly educated people.

7 Conclusion

A huge emphasis has been put on education policies in developing countries in the last ten years, especially thanks to the Millennium Development Goals. In India schooling enrollment rate has now almost reached 100%. There is no more need to prove the beneficial impact of education policies on income and growth. However, the way these education policies should be designed, in terms of which education level should be promoted, is still a matter of debate.

In this paper, my goal is to look at the relation between education inequality and income per capita in Indian States. I use data from 6 national sample survey rounds between 1983 and 2009-2010. As the Gini of education and the mean education level are highly correlated in each State, I use a methodology proposed by Berthélemy (2006) to separate in the Gini the level effect from the concentration effect. Using fixed-effects and System GMM to estimate the relation, I find that equality of education is negatively related to income per capita and that the rela-

tion is stronger in richer States. This result is robust to the use of a Theil index to measure the distribution of education and to the addition of a consumption Gini. When exploring the channels, I find that the three channels (non-linear returns of education, externalities and complementarities between workers) may be at stake in explaining the negative relation between equality of education and income per capita.

These findings underline important concerns. They show that the way an education system is designed has an impact on the income per capita of a country. Namely, this paper shows that at a given period of time, an economy with an important proportion of highly educated individuals is associated with higher income per capita than an economy with less educated individuals. I also find that the marginal impact of higher education is bigger than the one of secondary education. Should we conclude that all developing countries should stop investing in primary education to focus on higher education? The answer is no, for several reasons. First, there are equity concerns: investing only in higher education would favor those who are able to afford primary education. Second, there are dynamic issues. Highly educated individuals have been first in primary and secondary school. Therefore, it is important to invest in primary education in order to have people with higher education in the future. Third, the fact that the relation between equality of education and income per capita is not linear also underlines that policy makers have to be cautious in their design. The way education should be optimally distributed depends on initial conditions. One policy recommendation coming out of this chapter would be to instead focus on drop-out after primary school, in order to enhance secondary education investments.

This paper also opens the way to further research. In particular, it is important to explore additional non-linearities. The relation between the distribution of education and income per capita may for example depend on the production structure of the economy. In addition to have policy relevance, really understanding the relation between distribution of education and income per capita could help clarify the different results found in the literature.

Chapter 2

Education spillovers in farm productivity: Empirical evidence from rural India

1 Introduction

The business of private schools is flourishing in India. Between 2005 and 2012, the percentage of 6 to 14 year olds enrolled in private schools rose from 16.3% to 28.3%.[1] These private schools allow for a quick expansion of the education supply, in particular in rural India where the shortage of public schools is important. Indeed, private schools play a compensation role when the State cannot keep up with the fast increase in enrollment. Muralidharan and Kremer (2006) report that poorer States have more private schools per capita than public schools. Apart from filling the gap left by the State, private schools are also preferred by households because they are more effective than public schools in providing knowledge to students. The reasons given for these better results are that the student to teacher ratio is lower in private schools than in public schools and that teachers' absenteeism is less frequent (Kingdon, 1996; Muralidharan and Kremer, 2006).

This rapid take-off of private schools in India does not go unnoticed and is at the heart of Indian political debates. Most of the discussions are related to this better efficiency and how to provide access to these schools to poor households. However, one important but little debated question is the one of social returns to education. Indeed, if social returns are higher than private ones because of externalities for example, private financing of education is not optimal because individuals invest less in education than what would maximize the welfare of the society.

This paper investigates the existence of social returns to education in India by studying education externalities in agriculture in rural India. Although the rural focus of this study can seem very limited, this geographical restriction is actually relevant for a country like India where almost 70% of the population still lives in rural areas, among which 72% depends on agriculture.[2] Agriculture also accounts for 17% of the GDP, and employs 51% of the total labor force.[3]

Education externalities are not a new idea in the literature (Marshall, 1890; Lucas, 1988). However, because of the identification challenges that they entail, their empirical assessment is quite recent. Theoretically, there are several potential channels leading human capital to have a higher social than private return. The most evident is that an increase in the aggregate human capital of a defined geographical area can increase productivity above the direct effect of human capital on individual productivity (Moretti, 2004b). It can be due to learning spillovers, if low-skilled

[1] ASER reports 2005 and 2012. This percentage is calculated over all children, enrolled and out-of school. However the enrollment rate has been constant over the period (around 96%).

[2] Census of India 2011.

[3] Source: World Bank, data are from 2010. By way of comparison, agriculture represented 1.8% of total GDP in France in 2009 and 1.2% in the US in 2011.

workers learn from high-skilled workers, as Martins and Jin (2010) observe from portuguese firms for example. It can also be due to complementarity effects, as in the O-Ring theory of Kremer (1993) where the human capital of a worker has a marginal return which grows with the human capital of the other workers. Education also has positive externalities if it reduces the probability of getting involved in activities which produce negative externalities such as crime. Another indirect impact of education is through elections, if a more educated electorate takes better decisions on economy-related issues (Moretti, 2004a).

This paper focuses on the direct impact of education on income, by looking at the impact of neighbors' mean education on households' farm productivity. Because of the small scale of agricultural production in India, the channel through which neighbors could have an impact on farm productivity is the one of learning spillovers. As farmers are not working together on each others farm, there is no room for complementarity as in a firm. But more precisely, what kind of learning can take place? First, farmers can learn from their neighbors about agricultural technology. The fact that there are learning spillovers when there is a technological change has been widely asserted in the literature (see for example in the case of agriculture in developing countries Foster and Rosenzweig, 1995; Munshi, 2004; Conley and Udry, 2010). Neighbors' education can therefore influence households' *adoption* of technologies, through their own adoption of technology, which is related to their education level. It can also have a positive impact on the *use* of technologies, or more generally, neighbors' education can increase efficiency in the use of inputs (Weir and Knight, 2007). It can be the case if all farmers are not on the production frontier, which would be a signal that certain inputs, such as fertilizers, manure or grains are over or underused. Finally, neighbors can have an *allocative effect*: their education level would have an impact on farmers through a learning on which input or output to choose, given their relative prices (Kumbhakar and Lovell, 2003).

The existing empirical literature mainly focuses on education externalities in cities or in firms in developed countries (for the most cited contributions, see Acemoglu and Angrist, 1999; Moretti, 2004c,b). To my knowledge, very few papers study education externalities in agriculture in developing countries, and they suffer from limitations inherent to the study of peer effects, namely the problem of definition of peers, and the identification challenge due to group level unobservables. Appleton and Balihuta (1996), in a first attempt to take into account the impact of neighbors' education on agricultural productivity, introduce neighbors' average level of education in the production function of farmers in rural Uganda. They find that the proportion of farmers with a primary schooling is significantly related to

farmers' productivity. However, the fact that neighbors' education is calculated at the community level does not allow for the control of omitted community effects. Weir and Knight (2007) estimate average and stochastic production frontier with neighbors' education as a control variable in rural Ethiopia. Here, neighbors are defined as groups of households within "sites", so they are able to control for site dummies. They also find a positive relation between neighbors' education and average production, but they do not find any impact of neighbors' education on farmers' efficiency. Again, although some unobservables are controlled for, the strategy does not allow to be conclusive on the causality of neighbors' education impact. Asadullah and Rahman (2009) use the same methodology to study external returns of education in agriculture in Bangladesh. They control for village fixed effects while defining neighbors at a smaller level called "Bari".[4] On the contrary to the two previous papers, they find no evidence of external returns of education on farm productivity. However, the (absence of) results may be driven by the sample design: their database only reports information for two households by Bari. This definition of neighbors allows for village dummies but may be too restrictive to capture any external effect. These three papers underline the trade-off existing in the empirical literature: on one hand, it is necessary to control for geographical fixed effects to prove that the captured effect does not reflect a spurious correlation, but on the other hand, controlling for fixed-effects may lead to too restrictive definitions of neighbors.

The strategy in this paper follows Munshi and Myaux (2006) who exploit the social structure of rural Bangladesh to study the impact of other women's behavior towards contraceptive on own adoption of contraceptives. They take advantage of the fact that women interact solely within their religious group to define women's reference group and to rule out that the results are driven by group unobservables.

Here, I exploit a similar context, where social interactions are happening within castes. Although social mobility and social mixing have increased in urban India, life in villages is still organized along caste lines. Education spillovers are therefore expected to occur within caste, while cross-caste effects should be absent. I test this prediction using data from the round 2006 of the ARIS-REDS data, a household survey conducted in rural India. It provides detailed information about agricultural output for 4611 households in 226 villages distributed across 17 States.

I find that the education level of households from the same caste has a strong impact on farm productivity, but there is no cross-castes effects, and no effects from members from the same caste who do not have agriculture as their main activity. This pattern of results rules out that the estimated education spillover effect is driven

[4]Each village is composed of several neighborhood called "Baris".

by caste-level unobservables. Indeed, as shown by Munshi and Myaux (2006), and reexplained in section 3, these results can only be obtained without any education spillover effect if the omitted determinants of farm productivity are totally uncorrelated across castes, as well as within castes across occupations. This is not likely to be the case as it is argued in section 6.

This paper therefore provides consistent evidence that there is an impact of neighbors' education when neighbors are adequately defined, and this result is unlikely to be driven by omitted variables. It is however important to keep in mind that the identification relies on the hypothesis previously mentioned, that omitted variables correlated to groups' educational level and households' productivity are also correlated within villages and across occupations. If this is not the case, it is possible that the education level of members from the same caste proxies for other caste-level unobservable characteristics, and causality cannot be inferred.

The rest of the paper is organized as follows. Section 2 describes the theoretical framework and the empirical specification. Section 3 discusses the identification challenges and the empirical strategy used. Data and descriptive statistics are provided in section 3 and results in section 4. Finally a discussion on the identification hypotheses is provided in section 6 before concluding in section 5.

2 Education spillovers theoretically and empirically

The goal of this section is to quickly show how education spillovers can be theoretically formalized before describing my empirical specification.

2.1 Theoretical context

Households' agricultural output is produced according to a Cobb-Douglas production function, where productivity is allowed to depend on neighbors' education. Namely, output y of household i is produced with three factors of production, land l, labor n and physical capital k, according to the following specification:

$$y_i = A_i k_i^{\alpha} n_i^{\beta} l_i^{\gamma} \tag{2.1}$$

where each household has a specific productivity A_i. The external effect of human capital can be captured by allowing A_i to depend on the surrounding human capital, as in Moretti (2004b). In other words, A_i can be written as

$$A_i = f(\bar{E}) \tag{2.2}$$

where $\bar{E}$ is the average human capital in the neighborhood of the household farm. It is important to note that this specification assumes that neighbors' education augments the productivity of the three factors of production. Whereas other assumptions are possible, neighbors' education could only increase labor productivity for example, Moretti (2004b) underlines that it is empirically hard to distinguish between all alternative explanations. Moreover, it is likely that in the particular context of agriculture, education spillovers actually increase the productivity of the three factors of production.

2.2 Empirical specification

The estimation procedure is in two steps. Households' productivity is first estimated using a Cobb-Douglas production function. The external impact of education on agricultural productivity is then estimated in a second step.[5]

2.2.1 Households' productivity

Upon taking logs, equation (2.1) reads:

$$lny_i = lnA_i + \alpha lnk_i + \beta lnn_i + \gamma lnl_i \tag{2.3}$$

which can be rewritten as

$$lny_i = \alpha lnk_i + \beta lnn_i + \gamma lnl_i + u_i \tag{2.4}$$

where $u_i = lnA_i$. Households' productivity (in log) lnA_i, is therefore the error term of equation (2.4) and can be calculated through the estimation of equation (2.4).

Given data availability, u_i is obtained from the estimation of the following equation:

$$lny_{is} = \theta + \alpha k_{is} + \beta lnn_{is} + \gamma lnl_{is} + D_s + u_{is} \tag{2.5}$$

[5]This two-steps procedure aims at avoiding issues related to the endogeneity of factors of production or to the production functional form. While the estimation of households' productivity in the first step is most likely to be biased for the aforementioned reasons, the estimates of the variable of interest should not be affected because productivity is afterwards used as a dependent variable (Wooldridge, 2003). Robustness checks with an estimation in one step of the main results are presented in the appendix.

I allow for a constant θ. y_{is} is the farm production of household i in state s, k_{is} is a dummy variable which is equal to one if the household owns any agricultural equipment, n_{is} is the number of days worked on the land, l_{is} is the amount of land cultivated and D_s are state dummies which take into account the heterogeneity of productivity across states.

2.2.2 External effect of education

In a second stage the external effect of education δ is estimated by regressing the residuals from the estimation of equation (2.4) on neighbors' education. Neighbors' education is indexed cvs because it is calculated over people from the same caste c in village v in state s. I also allow the aggregate productivity to depend on households characteristics X_{icvs}, caste dummies C_c, state dummies d_s and a constant Θ.

$$\hat{u}_{icvs} = \Theta + \delta E_{cvs} + X_{icvs}\rho + C_c + d_s + v_{icvs} \tag{2.6}$$

3 Identification strategy

Identifying peer effects is a well known empirical challenge (see for example Manski, 1993; Brock and Durlauf, 2001). The three main issues are the endogeneity of network formation, the problem of correlated effects and the reflection problem (Manski, 1993). In section 3.1 and 3.2, I explain how I am addressing the two first issues. The reflection problem is not of concern here. It refers to the fact that if your neighbor influences you, then it is most likely that you also influence your neighbor. But the problem only arises when the group variable considered is a choice variable. It would be problematic for example if I were to look at the impact of neighbors' use of high-yielding grains on households' use of high-yielding grains. But here my variable of interest is neighbors' education level, which is predetermined compared to households' output.

3.1 Definition of neighbors

The study of peer effects suffers from one important issue which is the endogeneity of network formation (Jackson, 2008). It refers to the problem arising from the fact that people may choose the people with who they interact according to specific characteristics related to the output. In this context for example, if households with a high productivity choose to interact with households with a high education level, the estimate on peers' education level will be upwardly biased.

The problem of the endogeneity of network formation is here avoided by using a definition of peers which is exogenously determined. Using an exogenous definition of the peer group is a pretty common practice in the literature. However, because of the lack of precise data, the chosen definition is often geographical, such as the village (as in Case, 1992; Foster and Rosenzweig, 1995; Munshi, 2004, for example) or the neighborhood (as in Weir and Knight, 2007; Asadullah and Rahman, 2009). Here, I exploit the social structure of India to define peers, the caste group. As villages in the sample are not situated close to each other, what I call "neighbors" in the rest of the paper are members from the same caste group in the same village.

Caste is defined as an "hereditary, endogamous, usually localized group, having a traditional association with an occupation, and a particular position in the local hierarchy of castes. Relations between castes are governed, among other things by the concepts of pollution and purity, and generally maximum commensality[6] occurs within the caste" (Srinivas, 1962).[7] Although customs and traditions evolve quickly in India, these changes are mostly happening in urban India. In rural India, caste is still the group of reference within which interactions occur. Moreover, within villages, households are often clustered by castes. This characteristics of Indian villages accentuates the concentration of interactions within castes.

Here, households are grouped into four caste groups: the Scheduled Castes (hereafter SC), the Scheduled Tribes (ST), the Other Backward Classes (OBC) and the High Castes. The SC and ST are at the bottom of the hierarchy. The OBC are slightly higher in the hierarchy, but are still considered as low castes. The High Castes as their name indicate are at the top of the hierarchy. Although each of these broad caste groups is composed of several subcaste groups (or jatis), this is on the higher level of grouping that I focus. Indeed, when commitment or trust is an issue, or when social norms are being established, the relevant group is the subcaste group (Munshi and Rosenzweig, 2009; Munshi and Myaux, 2006). But in the context of learning, information flows seem to be happening within the broad social group. In the survey that I am using here, individuals were asked to identify the five most reliable persons in the village that they contacted to have information on various topics. Overall, approximately 70% of the persons mentioned are from the same broad caste group as the respondent, and 50% are from the same subcaste group. When focusing only on information flows related to agriculture, the proportions are similar. The results in the rest of this paper further support this definition of neighbors.

This definition of "neighbors" is of course an approximation of real interactions.

[6]That is to say the act of eating together.

[7]Additional information on castes can be obtained in Deliège (2004).

It may be that there are households from the same caste who do not influence each other or households from another caste group who have an impact on productivity. Not including in the neighbor's group households who have an impact or including households who have no impact is similar to a measurement error. But under the hypothesis that the measurement error is uncorrelated to the regressors, the estimates should be downward biased (Wooldridge, 2003).

3.2 The problem of group specific omitted variables

The most important problem in this paper is the problem of group specific omitted variables, also called correlated effects in the literature (Manski, 1993). This problem refers to the situation where the peers variable that we are interested in captures other group level variables that we cannot control for. The identification strategy that I use to deal with this issue is taken from Munshi and Myaux (2006). In their paper, they study the impact of the use of contraception of women from the same community on own use of contraception. They find that the coefficient on the use of contraception of women from the same community is positive and significant, whereas women from the other community in the village have no impact. They argue that this result could only reflect a spurious correlation between unobservables at the community level and contraceptive behavior if unobservables are fully uncorrelated across communities in the same village. In this section, I reproduce their argumentation and I apply it to my own specification. I am therefore closely following their paper.

I first illustrate the problem of omitted variables, by looking at the equation that I estimate. For purposes of clarity, the specification is simplified compared to equation (2.6). In particular, I do not allow for caste dummies or for state dummies, which allows me to drop the v and the s index:

$$\hat{u}_{ic} = \Theta + \delta_c E_c + \delta_o E_o + X_{ic}\rho + U_c + v_{ic} \tag{2.7}$$

where $\hat{u}_{ic}$ is household's agricultural productivity, E_c is the average of caste members' education in the same village, X_{ic} is individual characteristics and U_c represents the caste level characteristics that we cannot control for in the estimation, because U_c is unobserved. I allow for cross-caste effects by introducing E_o, which is the mean education level of village members from other castes. Because interactions are happening within castes, we expect that the education level of households from the same caste has a positive and significant impact, whereas the education level of households from other castes in the village should have no impact. In other words, we expect $\delta_c > 0$ and $\delta_o = 0$.

It is easy to see that if there is no impact of neighbors' education on agricultural productivity, the true specification is:

$$\hat{u}_{ic} = \Theta + X_{ic}\rho + U_c + v_{ic} \tag{2.8}$$

In this case, E_c in equation (2.7) proxies for the unobserved U_c if E_c and U_c are correlated. Put differently, a positive estimate of δ_c when the true value is zero could be obtained if the education level of people from the same caste in the village is correlated to unobserved caste characteristics that have an impact on agricultural productivity. As E_c cannot by itself perfectly proxy for U_c, in a model where there is no education spillovers E_o is an additional proxy for U_c.

Can we get $\hat{\delta}_c > 0$ and $\hat{\delta}_o = 0$, the expected result from a model with education spillovers, in a model without education spillovers? Yes, but only if E_o does not provide any information on U_c. This could only be the case if the unobserved characteristics of members from the same caste and members from other castes in the village, respectively U_c and U_o, are orthogonal. Therefore, to explain the pattern where $\hat{\delta}_c > 0$ and $\hat{\delta}_o = 0$ *without education spillovers*, U_c and U_o must be totally uncorrelated within the village. When estimating equation (2.7), getting $\hat{\delta}_c > 0$ and $\hat{\delta}_o = 0$ consequently rules out that E_c, the education level of households from the same caste, captures caste level omitted variables *correlated across castes* in the village.

However, it does not rule out the case where E_c captures omitted variables which are uncorrelated across caste groups. Even if it is hard to think about unobserved variables correlated to the education level of one caste group and to agricultural productivity but not correlated to the education level of other households in the village as I argue in section 6, this situation should not be excluded. To take that into account, I now estimate the impact of households from the same caste who have agriculture as their main activity, whose education level is E_{AGRc}. The other group is members from the same caste who do not have agriculture as their main activity, and their education level is noted $E_{nonAGRc}$. The equation to estimate is:

$$\hat{u}_{ic} = \Theta + \delta_{AGRc}E_{AGRc} + \delta_{nonAGRc}E_{nonAGRc} + X_{ic}\rho + U_{AGRc} + v_{ic} \tag{2.9}$$

The same reasoning can be applied here. If $\hat{\delta}_{AGRc} > 0$ and $\hat{\delta}_{nonAGRc} = 0$, then we can rule out that E_{AGRc} captures the unobserved characteristics U_{AGRc}, if U_{AGRc} is not totally uncorrelated with the unobserved characteristics of households from the same caste but not cultivating land $U_{nonAGRc}$. Given the social structure of India, it is really unlikely that there are omitted variables completely uncorrelated within castes. This point is discussed in section 6.

4 Data and descriptive statistics

4.1 Data and variables definition

The data used to conduct this study are from the 2006 round of the ARIS-REDS database from the National Council of Applied Economic Research (NCAER). Since 1971, the NCAER has been conducting surveys on a *sample* of households in 232 villages in the 17 major States of India. The round of 2006 has the advantage to also have a *census* of every household in each village with basic information on households and households' heads such as their demographic characteristics (gender, age, education level, size of the household) and their main occupation. As detailed information about agriculture is not provided in the census, I have to restrict my analysis to households in the *sample* who cultivate land. The identification strategy also requires several castes per villages and that in each caste, there are households who do not have agriculture as their main activity. The final sample is therefore composed of 1611 households in 226 villages. Neighbors' education is calculated using the census.

The first set of variables are the agriculture related variables, used to estimate the production function. The dependent variable is households' total agricultural output.[8] Most households cultivate several crops, so total output cannot be calculated using quantities. I aggregate the different crops using their value in rupees, declared by the household. When the production function is estimated separately for wheat and rice, total agricultural output is a quantity, measured in quintals.[9] The factors of production are the area of land cultivated (in acres), the number of days worked on the land (this measure includes family workers as well as hired labour), and variables proxying for the capital used, namely mechanized assets and non-mechanized assets. Mechanized assets and non-mechanized assets are dummy variables equal to one if the household respectively owns at least one mechanized,[10] or non-mechanized[11] asset. I also control for the proportion of irrigated land and the proportion of land owned (as opposed to rented in) by the household over total

[8] I only focus on field crops. Plantation crops, such as coffee or tea are excluded, to keep some homogeneity in my sample. It is however not a strong restriction because only six households in the data cultivate plantation crops.

[9] One quintal is 100 kg.

[10] Mechanized assets include tractors, trailers, threshers, electrified motors, non-electrified motors (oil engine), gauge wheels, ploughs, tractors with trailer, tractors with thresher, tractors with oil engine, tractors with gauge wheel, tractors with plough, disc harrows, tillers or cultivators, plough discs or mould boards seed drills, power tillers, power sprayers, chaff cutters, cane crushers, combine harvesters.

[11] Non-mechanized assets include iron ploughs, cultivators, harrows, levelers, hoes or manual earth removers, seed rills, weeders, sprayers or dusters, winnowers or pitch forks, bullock carts, chaff cutters, sickles, scissors.

land cultivated. When the productivity is estimated separately for wheat and rice, because I do not have the precise information per crop, the variable for irrigation is just a dummy equal to one if the land on which the crop is cultivated is irrigated, and the variable for land ownership is also a dummy variable equal to one if the land is owned by the household.

The second set of variables used is household level demographic characteristics, namely the age of the household head, her/his gender, her/his education level (in years of schooling) and the number of members in the household. There are also dummies indicating the caste group of the household. Neighbors' education level, depending on the specification, is the mean education level of households' heads from the same caste in the village or the mean education level of households' heads from the same caste who have agriculture as their main occupation. In both cases, the education level of the household's head is excluded from the mean calculation.

4.2 Descriptive statistics

Table 4.2 reports descriptive statistics for the whole sample of households having agricultural output. In mean, the head of the household is around fifty years old and has primary education (5 years of education). Only 5% of the households' heads are women and a household is composed on average of 6 people. In terms of agricultural equipment, only a third of the households have a mechanized asset but almost every household has at least one non-mechanized asset (96%). Interestingly, a big proportion of land which is cultivated is owned by the household and is irrigated. Also, new technologies are widely used: 63% of the households have at least one crop for which they are using high yielding varieties (thereafter HYV).

However, when we look at the statistics separately by caste, the situation is very heterogenous across caste groups. Although pure demographic characteristics such as age, household size, and percentage of households with a female head do not differ strongly from one group to the other, the other variables such as education or cultivated areas are very different and follow the traditional hierarchy. Scheduled castes and scheduled tribes have a lower education level and cultivate less land than OBC, who are themselves worse-off than high castes. Similarly the proportion of low caste land which is irrigated or the probability that a low caste has mechanized equipment is much lower than for a high caste. The fact that human capital and wealth is distributed along caste lines is not very surprising. It is a widely reported phenomenon in the literature (see e.g. Kijima, 2006). These statistics confirm that in rural India, the economic situation is still highly related to the caste one is born in.

I now turn to the descriptive statistics for neighbors. As a reminder, neighbors'

Table 2.1: Descriptive statistics

	Full sample		SC		ST		OBC		High castes	
Households' demographic characteristics										
Age (Years)	50.81	(13.15)	50.47	(13.12)	47.42	(12.66)	50.49	(12.75)	52.36	(13.68)
Household Size	6.10	(3.28)	5.71	(2.78)	5.85	(3.46)	6.23	(3.21)	6.11	(3.50)
Female head	0.05	(0.23)	0.06	(0.24)	0.06	(0.23)	0.05	(0.22)	0.05	(0.23)
Head education level	5.13	(4.91)	4.36	(4.59)	2.84	(3.83)	5.02	(4.90)	6.21	(5.01)
Farms' characteristics										
Cultivated area (in acres)	6.92	(10.08)	3.51	(4.06)	5.98	(6.50)	6.62	(9.06)	8.96	(13.11)
Days worked on field	130	(140)	95	(99)	92	(73)	131	(145)	151	(154)
Mechanized assets	0.33	(0.47)	0.10	(0.31)	0.15	(0.36)	0.32	(0.47)	0.48	(0.50)
Non-mechanized assets	0.96	(0.19)	0.96	(0.19)	0.98	(0.13)	0.96	(0.19)	0.95	(0.21)
Proportion of land owned	0.93	(0.24)	0.87	(0.32)	0.98	(0.13)	0.92	(0.25)	0.94	(0.21)
Proportion of irrigated land	0.64	(0.46)	0.54	(0.49)	0.44	(0.47)	0.68	(0.45)	0.68	(0.45)
HYV	0.63	(0.48)	0.58	(0.49)	0.64	(0.48)	0.58	(0.49)	0.73	(0.45)
Neighbors characteristics										
All neighbors										
Mean age	47.99	(3.48)	47.45	(3.53)	45.38	(2.78)	47.54	(2.78)	49.60	(3.88)
Mean HH size	5.51	(1.01)	5.25	(0.82)	5.25	(0.86)	5.59	(1.03)	5.57	(1.04)
Female head	0.09	(0.05)	0.10	(0.06)	0.09	(0.06)	0.08	(0.05)	0.09	(0.06)
Mean education level	4.61	(2.04)	3.75	(1.73)	2.71	(1.97)	4.34	(1.71)	5.86	(1.92)
Neighbors cultivating land										
Mean age	49.46	(3.90)	50.60	(4.43)	46.82	(4.01)	49.16	(3.24)	50.20	(4.19)
Mean HH size	5.80	(1.12)	5.57	(1.10)	5.57	(0.97)	5.94	(1.13)	5.73	(1.11)
Female head	0.06	(0.06)	0.07	(0.08)	0.07	(0.08)	0.06	(0.05)	0.06	(0.06)
Mean education level	4.78	(2.23)	3.94	(2.16)	2.74	(2.04)	4.59	(2.02)	5.94	(1.97)
Observations	4611		556		387		2226		1442	

The table reports the means for each variable and standard deviations are in parentheses.

characteristics are calculated with the data that provide information for every household in the village (what is called the *census*). The first set of neighbors' descriptive statistics is when neighbors are defined as members from the same caste, whatever the occupation. We can see that neighbors are slightly different than the sample of farmers. They have smaller households, and a lower education level in mean. Reassuringly, when we only look at neighbors with farming as their main activity, neighbors are really similar to the sample of farmers. The only difference is on the education level that is lower for farmers. It may be due to the fact that the definition of farmers is different in the *sample* and in the *census*. While in the sample a farmer is anybody who has some agricultural output, in the sample farmers are only those who have agriculture as their main activity.

5 Results

5.1 Productivity estimation

The first step, before estimating the impact of neighbors' education, is to get the farm productivity of each household. To do that, I estimate equation (2.5) and take the residuals. The results from the estimation is reported in column 1 of table 2.2. The subsequent columns show the estimation of the productivity equation for wheat (column 2) and for rice (column 3). The samples in those columns are therefore restricted to households cultivating wheat and rice respectively.

The results are very similar in the three columns. The R^2 is very high: the variables explain up to 85% of the variation in output. Not surprisingly, land is the most important factor of production, followed by labor. Having irrigation also seems to matter a lot. The reason why the coefficient on irrigation is very different once the estimation is made separately for wheat and rice is because the variable is not measured the same way.[12] Finally, owning the land which is cultivated instead of renting does not seem to matter for productivity.

5.2 Caste and cross-caste impact

The residuals of equation (2.5) are now used as a measure of households' productivity. I look at the impact of the education of members from the same caste and of the education of the other castes in the village. All the regressions are estimated with state dummies. The first set of results is in table 3 column 1 to 3. In column 1, the only control variables are the household demographic characteristics. In column 2, I add control variables at the caste in the village level, namely the number of

[12]See section 3 for an explanation on the variables construction.

Table 2.2: Productivity estimation

Dependant variable:	Total agricultural output		
	(1) All	(2) Wheat	(3) Rice
Cultivated area (in log)	0.765***	0.802***	0.817***
	(0.0158)	(0.0217)	(0.0222)
Days worked (in log)	0.252***	0.140***	0.168***
	(0.0164)	(0.0222)	(0.0224)
Mechanized assets	0.158***	0.0749***	0.0674***
	(0.0193)	(0.0229)	(0.0213)
Non-mechanized assets	-0.0363	0.131	-0.140***
	(0.0450)	(0.0940)	(0.0458)
Land owned	-0.00752	0.0154	-0.0116
	(0.0378)	(0.0461)	(0.0370)
Irrigated land	0.243***	0.0501**	0.0165
	(0.0249)	(0.0227)	(0.0241)
State dummies	yes	yes	yes
N	4611	2182	2398
r2	0.806	0.838	0.851

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are reported in parentheses. The dependent variable in column 1 is total production, in column 2 it is wheat production and in column 3 it is rice production. Some people neither cultivate wheat nor rice, so the number of observations in column 2 and 3 do not sum up to the number of observations in column 1.

caste members in the village and the caste mean value of households characteristics (mean age of households' heads, percentage of female heads in the caste, and mean household size). In column 3, I additionally control for the same variables calculated over members of the other castes in the village. The results are very stable across specifications and robust to the addition of caste level controls. I find that the mean level of education of members from the same caste in the same village is strongly positive and significant: an additional year in the mean level of education of castemates increases household productivity between 4.6 and 5.9%. On the contrary, the education level of members from other castes in the village has no impact. In the three specifications, the coefficient is close to zero and not significant.

For the other variables, the only household's characteristic which has an impact on productivity is households' size. For a given quantity of labor, having more family members increases productivity. It can be due to hired labor being less productive than family members, which is in line with what Rosenzweig and Foster (2010) find for India. Surprisingly, the head's education has no impact. Although it can be a

Table 2.3: Within villages: caste and cross-caste effect

Dependant variable:	Household Productivity				Placebo
	(1)	(2)	(3)	(4)	(5)
Age (Years)	0.000872	0.000928	0.000988	0.000999	668.4***
	(0.000651)	(0.000633)	(0.000620)	(0.000620)	(155.7)
Female head	0.00280	0.00147	-0.00226	-0.00728	1729.7
	(0.0370)	(0.0357)	(0.0343)	(0.0347)	(3272.1)
Household Size	0.00742**	0.00693***	0.00768***	0.00774***	1494.1**
	(0.00287)	(0.00247)	(0.00248)	(0.00247)	(707.5)
Head education level	0.00146	0.00135	0.00198	0.0168**	1913.9***
	(0.00202)	(0.00199)	(0.00201)	(0.00759)	(394.4)
Scheduled Caste	0.0487	0.0648	0.100	0.104	-21141.7**
	(0.0861)	(0.0876)	(0.0848)	(0.0841)	(8485.3)
Scheduled Tribe	0.197**	0.210**	0.201**	0.219**	-44849.7**
	(0.0997)	(0.0955)	(0.0910)	(0.0917)	(17569.8)
Other Backward	0.0977*	0.0942*	0.109*	0.109*	-21961.9**
Classes	(0.0564)	(0.0566)	(0.0558)	(0.0560)	(10973.1)
Caste mean education	0.0509***	0.0592***	0.0461**	0.0646***	-3391.1
level	(0.0166)	(0.0183)	(0.0183)	(0.0222)	(2369.1)
Head educ * Caste				-0.00308**	
mean education				(0.00149)	
Other castes mean	0.00159	0.0000894	0.00463	0.00258	1988.0
education level	(0.0153)	(0.0154)	(0.0141)	(0.0141)	(1350.8)
State dummies	yes	yes	yes	yes	yes
Caste controls	no	yes	yes	yes	yes
Other caste controls	no	no	yes	yes	yes
N	4611	4611	4611	4611	4604
r2	0.0359	0.0442	0.0728	0.0761	0.236
Head educ & interact				0.0826	
Interact & Caste educ				0.0138	
Impact for caste educ					
Head educ: 0 years				0.0646***	
Head educ: 5 years				0.0492***	
Head educ: 10 years				0.0338*	
Head educ: 15 years				0.0184	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, corrected for clustering at the caste in the village level. Caste controls and other caste controls include the number of people in each group and the mean value for the demographic characteristics (age, proportion of female head, household size). P-values for the joint significance between the head education level and the interaction term and between caste mean education level and the interaction term are reported. The bottom of the table also reports the marginal impact of caste education for various values of the head education.

mechanical effect due the external effects of education,[13] an alternative explanation

[13]Weir and Knight (2007) underline that the effects of education may not be apparent if less

is that households' heads education and neighbors education are substitutes. Indeed, when the education of members from the same caste is not included (not reported here), the household's head education has a positive and significant impact. In column 4, I test this explanation by interacting neighbors' education with households head education. The interaction terms turns out to be negative, which underlines that neighbors' education and heads' education are substitute: the positive impact of neighbors' education decreases with households' heads education. For example, an additional year of education in castes' mean education level increases households' productivity when the head has no education by 6.5 %. However, when the head has a university diploma (15 years of education), neighbors do not matter anymore. Reassuringly, with this specification, education also has a private return.

So far, the results in table 2.3 have shown that there is a positive impact of neighbors' education on households' productivity. The fact that the education of other castes has no impact on households' productivity rules out that the estimated caste effect is driven by unobserved caste variables *correlated* across castes as explained in section 3.2. However, one can still think that the impact is due to unobserved caste variables *uncorrelated* across castes. For example, the fact that castes are geographically clustered in villages may lead to this kind of spurious correlation. It can be that schools settle in the most prosperous areas of villages, in which case castes situated in those areas have an easier access to education, as well as a higher productivity.

To test if this is this kind of unobservable uncorrelated across castes that is driving the impact, column 5 presents the results of a falsification test, where the dependent variable is households' savings. There is arguably no education spillovers effects in savings behavior, notably because the amount that households save is hardly observable by their neighbors. There is indeed empirical evidence that households in developing countries develop strategic behaviors to hide their savings, in order to avoid "taxation" from their network (Baland et al., 2011; Di Falco and Bulte, 2011). However, we can think that savings are correlated to other caste level variables, such as caste wealth.[14] Therefore, if the positive impact of neighbors' education on productivity was driven by unobservables at the caste level, then we would expect that the coefficient on neighbors' education in the falsification test is also positive and significant. But the results in column 5 confirm the interpretation of the positive coefficient on caste members education as an education spillover effect: the amount of savings is not affected by caste members' education.

educated farmers copy more educated farmers

[14]Insurance mechanisms have been observed to happen within castes (Munshi and Rosenzweig, 2009), which means that households savings may be affected by the other caste members income or wealth.

5.3 Within castes: same occupation and cross-occupation effect

The results in section 5.2 show a positive impact of caste members' education on households' productivity. The fact that the education level of households not from the same caste in the village has absolutely no impact and the falsification test underline that the results do not seem to be driven by caste level unobservables.

As explained in section 3.2, an alternative way to check if the measured spillovers comes from unobservables correlated to caste's education and productivity is to only focus on households from the same caste. Caste members are now divided between those who have agriculture as their main activity and the others. As underlined in the introduction, the channels leading to education spillovers in agriculture are related to agriculture so we do not expect education spillovers to go from households not having agriculture as their main occupation to households having agriculture as their main occupation. The interpretation of the results is as previously. If there is no cross-occupation impact, then we can rule out that the coefficient on neighbors' education reflects a spurious correlation due to unobserved characteristics uncorrelated across occupations within caste.

Table 2.4 shows the results. They are organized as in table 2.3: column 1 shows the main specification. Although I control for households' characteristics as in table 2.3, their coefficients are not reported because they do not provide additional information. Again the results are consistent with an education spillover effect: households from the same caste who have agriculture as their main activity have a strong and significant impact on households' productivity. An increase of one year in their mean education level increases household's productivity between 4.5 and 5.2%. On the contrary, the variable which measures the education level of neighbors from the same caste but who are not farmers has no impact: the coefficient is very close to 0 and not significant.

Columns 2 and 3 test the robustness of these results by adding additional caste-occupation level controls. The results are very robust. There is a strong impact from households who cultivate land whereas those who do not have no impact. Column 4 looks at the complementarity between head's education and neighbors education. Again, it seems that households' heads education and neighbors education are substitutable: the impact of neighbors' education diminishes with households' level of education. Finally column 5 runs the same falsification test as in table 2.3 with households' savings as the dependent variable. There is no impact of neighbors' education on savings. This result, along with the absence of cross-occupation impact reinforces the argument that the previously measured impact of neighbors' education is not driven by unobservables.

Table 2.4: Within caste: same occupation and cross-occupation effect

Dependant variable:	Household Productivity				Placebo
	(1)	(2)	(3)	(4)	(5)
Head education level	0.00178	0.00165	0.00147	0.0177***	1778.2***
	(0.00203)	(0.00198)	(0.00198)	(0.00685)	(393.7)
Same occup caste	0.0453***	0.0503***	0.0517***	0.0691***	-2289.0
mean education level	(0.0147)	(0.0151)	(0.0149)	(0.0171)	(2336.8)
Head educ * caste				-0.00321**	
agr mean education level				(0.00129)	
Other occup caste	-0.00107	-0.00114	-0.00177	-0.00147	1440.5
mean education level	(0.0143)	(0.0152)	(0.0150)	(0.0150)	(2233.1)
HH controls	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes
Caste occup controls	no	yes	yes	yes	yes
Caste oth occup controls	no	no	yes	yes	yes
N	4611	4611	4611	4611	4604
r2	0.0342	0.0402	0.0452	0.0494	0.222
Head educ & interact				0.0359	
Interact & Caste educ				0.0003	
Impact for caste educ					
Head educ: 0 years				0.0691***	
Head educ: 5 years				0.0530***	
Head educ: 10 years				0.0370**	
Head educ: 15 years				0.0210	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, corrected for clustering at the caste in the village level. Caste occup controls and caste oth occup controls include the number of people in each group and the mean value for the demographic characteristics (age, proportion of female head, household size). P-values for the joint significance between the head education level and the interaction term and between caste age mean education level and the interaction term are reported. The bottom of the table also reports the marginal impact of caste age education for various values of the head education.

5.4 Within jati: same occupation and cross-occupation effect

I now look at the impact of members from the same caste but not from the same subcaste (or jati). In section 3.1, the definition of households' group of reference was driven by the idea that social proximity was required for learning to happen and the results in the previous sections have confirmed that other castes group in the village do not matter. The choice of the broad caste group over the jati was also made because there is no commitment or trust issues in agricultural learning, and therefore learning is likely to cross jatis boundaries. In this section I test this last hypothesis by looking at the impact of members from the same jati and members from the same caste but not from the same jati. If the hypothesis is right, we expect

members from the same caste but not from the same jati to also have an impact on households' productivity.

Table 2.5: Within jati: same occupation and cross-occupation effect

Dependant variable:	Household Productivity				Placebo
	(1)	(2)	(3)	(4)	(5)
Head education level	0.00121	0.00124	0.00109	0.0202***	1818.8***
	(0.00223)	(0.00218)	(0.00218)	(0.00708)	(487.3)
Same occup jati mean	0.0306***	0.0311***	0.0344***	0.0566***	1567.3
education level	(0.0101)	(0.0105)	(0.0108)	(0.0154)	(1238.2)
Head educ * jati agr				-0.00378***	
mean education level				(0.00133)	
Other occup jati	0.0217*	0.0214	0.0276*	0.0256*	-633.5
mean educaiton level	(0.0128)	(0.0132)	(0.0141)	(0.0142)	(1074.1)
HH controls	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes
Jati occup controls	no	yes	yes	yes	yes
Caste occup oth jati	no	no	yes	yes	yes
N	3001	3001	3001	3001	2995
r2	0.0634	0.0647	0.0803	0.0868	0.188
Head educ & interact				0.0160	
Interact & Caste educ				0.0012	
Impact for caste educ					
Head educ: 0 years				0.0566***	
Head educ: 5 years				0.0376***	
Head educ: 10 years				0.0188*	
Head educ: 15 years				-0.0001	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, corrected for clustering at the jati in the village level. The sample is restricted to jatis that have at least three members. Jati occup controls and caste oth jati controls include the number of people in each group and the mean value for households' demographic characteristics (age, proportion of female head, household size). P-values for the joint significance between the head education level and the interaction term and between jati age mean education level and the interaction term are reported. The bottom of the table also reports the marginal impact of jatis' mean education for various values of the head education.

Table 2.5 shows the results. I only consider households who have agriculture as their main activity, but the results are similar when people not having agriculture as their main activity are included. As expected, the education of households from the same jati has a positive impact on agricultural productivity. One additional year in the educational level of households from the same jati increases households' productivity by 3%. But more interesting, households from the same caste but not from the same jati also have a positive and significant impact on households productivity. This result confirms that *economic* interactions are cutting jatis boundaries.

5.5 Who benefits from education spillovers? Heterogeneity of the impact

On average, there is an education spillover effect. The mean education level of households from the same caste and from the same occupation has a positive impact on agricultural productivity. This section explores if there are non-linearities in the impact. The first part studies if the impact differs across castes and the second part considers heterogeneity across crops.

5.5.1 Heterogeneity across castes

The heterogeneity of impact across castes is studied by adding an interaction term between the caste group from which the household is and the mean education level in this group. The estimations are reported in table 2.6 and the impact by caste is reported at the bottom of the table. Again, I only present the results where neighbors are defined as only people cultivating land, but when neighbors are defined more broadly the results are similar. The default group in the table is the group of OBC.

What we can see is that the impact does not differ much across castes, except for the high castes. For the SC, ST and OBC, the coefficient on the education level of households from the same caste is quite large. It is not significant for SC, but given the size of the coefficient it may just be a problem of number of observations.[15] However, for the high castes the coefficient is much smaller and not significantly different from zero. One explanation may lie in the fact that high castes have a higher productivity than other groups on average. They may therefore be closer to the productivity frontier and have less to learn from their neighbors than the other castes.

5.5.2 Heterogeneity across crops

Another source of heterogeneity may be in the type of crops cultivated. In section 5.2 and 5.3, I do not differentiate across crops, even if there are good reasons to believe that the impact differs. The extent to which learning can happen for a specific crop may depend on the difficulty to cultivate this crop, as well as the level of technology used for cultivation. In this section, I look at the impact of neighbors' education separately for two crops, rice (paddy) and wheat.

Rice and wheat are the two main crops cultivated in India. In my sample, 27% of the households only cultivate wheat, 31% only cultivate rice and 21% cultivate both. Therefore only 21% cultivate neither wheat nor rice. Additionally, except for the States on the western ocean side (Andhra Pradesh, Orissa, Tamil Nadu and West

[15] The coefficient is significantly different from zero at a 15% level.

Table 2.6: Heterogeneity of the impact across castes

Dependant variable:	Household Productivity		
	(1)	(2)	(3)
Years of schooling	0.00158	0.00154	0.00138
	(0.00202)	(0.00197)	(0.00197)
SC	-0.00354	0.0454	0.0387
	(0.187)	(0.196)	(0.196)
ST	0.111	0.143	0.141
	(0.171)	(0.169)	(0.162)
High Castes	0.172	0.159	0.149
	(0.136)	(0.137)	(0.136)
Same occup caste mean education level	0.0586***	0.0633***	0.0636***
	(0.0207)	(0.0213)	(0.0208)
SC * same occup caste mean educ	-0.00821	-0.0106	-0.0107
	(0.0398)	(0.0408)	(0.0410)
ST * same occup caste mean educ	0.00492	0.00231	0.00418
	(0.0327)	(0.0326)	(0.0316)
High Castes * same occup caste mean educ	-0.0454*	-0.0428*	-0.0395
	(0.0250)	(0.0251)	(0.0249)
Other occup caste mean education level	0.00205	0.00193	0.00121
	(0.0150)	(0.0159)	(0.0158)
HH controls	yes	yes	yes
State dummies	yes	yes	yes
Caste occup controls	no	yes	yes
Caste oth occup controls	no	no	yes
N	4611	4611	4611
r2	0.0398	0.0449	0.0492
Impact of caste educ			
SC	0.0504	0.0527	0.0529
ST	0.0636**	0.0656**	0.0678**
OBC	0.0586***	0.0633***	0.0636***
High castes	0.0132	0.0205	0.0242

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, corrected for clustering at the caste in the village level. Caste occup controls and caste oth occup controls include the number of people in each group and the mean value for the demographic characteristics (age, proportion of female head, household size). The bottom of the table reports the marginal impact of caste mean education for each caste group.

Bengal) and Kerala where there is no wheat cultivated, most of the States cultivate both crops.[16] These two crops have seen a huge increase in their productivity during Green Revolution which began at the beginning of the 70s, thanks to the diffusion of HYV. However, the process of adoption has been slower for rice, because the

[16]Rajasthan is an exception: in my sample only 2 households cultivate rice. Therefore it is excluded from the rice estimation.

productivity of the first generation of rice HYV was very dependent on local conditions. Munshi (2004) shows that learning from neighbors was lower for rice HYV than for wheat HYV during the Green Revolution. This finding is also consistent with the papers previously quoted on education spillovers, although the grain type is not given as a possible explanation of the differences in findings. Appleton and Balihuta (1996) and Weir and Knight (2007), study education spillovers respectively for agricultural productivity in general[17] and cereals productivity, and they do find that there is an impact. On the contrary, Asadullah and Rahman (2009) look at the impact of education spillovers on rice productivity and do not find any evidence that neighbors' education matter. One can think that it is due to the smaller learning potential in rice production.

Table 2.7 provides independent estimations for rice and wheat. As previously, the dependent variable is household's productivity, which is the residual taken from the separate estimation for rice and wheat of equation (2.5). In column 1 and 3, I report the results for the main specification. The impact of neighbors' education is indeed different for wheat and rice. For wheat, the impact is big and similar to what was previously measured. The impact is smaller for rice. The results are also different for the impact of the household head' education. For wheat, similarly to the aggregate productivity, there seems to be no impact of households' own education. On the contrary, for rice the education of households' heads is positive and significant. This result is in line with Munshi (2004), who finds that there is learning from neighbors for wheat, but that rice growers experiment more on their own land to compensate for the lack of learning.

However, in Munshi (2004), the absence of learning from neighbors for rice was specifically related to HYV adoption. The impact of neighbors' education should consequently differ depending on wether the farmer is using HYV or traditional grains. To test the hypothesis that the lower impact for rice is related to the technology used, in column 2 and 4, I interact neighbors' education with a dummy variable equal to one if the household is using HYV grains. For wheat, the impact of neighbors' education does not depend on the technology used. The interaction term between neighbors' education and using HYV grains is not significant and very close to zero. However, for rice, the results confirm that learning depends on the type of technology used. When households use traditional grains, neighbors have a strong and significant impact: a one year increase in neighbors' education increases rice productivity by 4.6%. But when the household is using HYV, the coefficient on neighbors' education is small and not significantly different from zero (bottom line of table 2.7). This result underlines that education spillovers are in mean smaller for

[17]The production of the different crops is aggregated.

Table 2.7: Impact for wheat and rice

Dependant variable:	Household Productivity			
	(1)	(2)	(3)	(4)
Crop	Wheat		Rice	
Head education level	0.00169	0.00135	0.00645***	0.00643***
	(0.00211)	(0.00206)	(0.00203)	(0.00202)
Same occup caste	0.0362***	0.0358**	0.0261**	0.0464***
mean education level	(0.0124)	(0.0148)	(0.0128)	(0.0146)
Other occup caste	0.00630	0.00876	-0.0107	-0.0128
mean education level	(0.0121)	(0.0118)	(0.0122)	(0.0120)
Neighbors' educ* HYV		-0.00306		-0.0316**
		(0.0133)		(0.0138)
HYV		-0.0589		0.244***
		(0.0792)		(0.0867)
HH controls	yes	yes	yes	yes
State dummies	yes	yes	yes	yes
Caste occup controls	yes	yes	yes	yes
Caste oth occup controls	yes	yes	yes	yes
N	2210	2207	2415	2408
r2	0.0569	0.0608	0.0557	0.0645
HYV & interact		0.0287		0.0160
Interact & Caste educ		0.0209		0.0052
Impact for caste educ				
HYV = 0		0.0358**		0.0464***
HYV = 1		0.0328 **		0.0147

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, corrected for clustering at the caste in the village level. Casteoccup controls and casteothoccup controls include the number of people in each group and the mean value for the demographic characteristics (age, proportion of female head, household size). P-values for the joint significance between cultivating HYV and the interaction term and between caste age mean education level and the interaction term are reported. The bottom of the table also reports the marginal impact of caste age education when grains are HYV and when they are not.

households who produce rice because there is no, or very little learning happening for households using rice HYV grains.

6 Discussion

Section 4 shows that the mean education level of households from the same caste and having agriculture as their main activity is positively correlated to households' farm productivity. This result can be interpreted as causal if there is no group level unobservables affecting both households' productivity and castes' mean level of education. Moreover, given that households not from the same caste and households

from the same caste but who do not have agriculture as their main activity have no impact on households' productivity, even if there are caste level unobservables, the relation is not causal only if those unobservables are fully uncorrelated across castes, and within castes across occupations, as explained in section 3.2. In this section I discuss the likelihood that castes' mean education level is correlated to caste level unobservables having an independent impact on productivity, and consequently the likelihood that this variable captures other effects that education spillovers.

The mean education level of a caste in an area depends on the supply of education, as well as the demand for education of this caste. It is very unlikely that the results are driven by a supply side story. Indeed, getting the results of table 2.3 without spillovers effects would require that the unobservables that are correlated to castes mean education level and households' productivity are totally uncorrelated across castes. All the variables that one can think of that would have an impact on the supply of education, are correlated within villages. For example, the presence of a school in the village, the distance to the closest city or schooling quality are correlated across castes within villages. Desai and Kulkarni (2008) however report that in some villages, scheduled castes are not allowed to cross higher castes areas and consequently that dalit children have a very long way to go to school, because they have to walk along the periphery of the village. This kind of discrimination could create artificial differences in the education supply for different castes within a village. But even if this is the case, it is hard to imagine that the supply of education is fully uncorrelated within castes. There is no particular reason why a scheduled caste child whose family cultivates land has a better access to schooling than a scheduled caste child whose family has another occupation. No supply side explanations could explain the pattern of results that I get in table 2.4.

On the demand side, the situation is more complex. Demand for education depends on several things, such as ability, wealth of the family, or expected returns to education. Whereas there are good reasons to believe that ability is randomly distributed across castes and occupations, wealth and returns to education may be more problematic. It is well reported that wealth is still distributed along castes line in India, and that low castes have a lower return to education (Kijima, 2006). It is possible to imagine that wealth and returns to education are totally uncorrelated across castes in a village.

If the unobservables are systematically making low castes worse off than high castes, say for example that low castes have systematically the least fertile lands and therefore invest less in education than higher castes, then it is controlled for

by the caste dummies. In particular, the caste dummies control for the hierarchy across castes, which is at the origin of the distribution of wealth and opportunities along castes lines.

However, it can be that unobservables impacting agricultural productivity, and therefore wealth, are uncorrelated across castes but not in the same way in every village. One can think about three situations that would give the same results as in section 4 without any education spillover. First, the *quality* of land can differ across castes in a village. We could get a spurious correlation if castes in less fertile areas invest less in education, because they are poorer due to the low fertility of their land. In the same way, the *suitability* of land for crops could vary across castes in the village. This would lead different castes in the village to cultivate crops with different yield. Finally, castes' agricultural productivity can depend on their relative social status in the village. Anderson (2011) for example shows that in villages in North India that are dominated by high castes, the productivity of low castes land is lower than in villages dominated by low castes. She shows that it is due to a trade breakdown in irrigation water across caste groups. These three potential sources of variation in land productivity across caste groups could in turn affect wealth and educational investments.

Regarding land quality, it is not highly probable that it is fully uncorrelated within a same village. Most of the variation of land quality is from one village to the other. The choice of crop could be a more serious source of bias. However, in India, villages are pretty homogeneous in the type of crop that are cultivated. On average, the same crop is cultivated on 60% of the land in the village, and less than 7% of the households do not cultivate this crop. Moreover, the proportion of households that do not cultivate the main crop is broadly the same in each caste group. The fact that the effect of caste members' education holds when only looking at rice growers and wheat growers also helps ruling out the idea that the measured impact is driven by crop choice. I test the last alternative explanation, that the social structure of the village matter for agricultural productivity, by adding a dummy variable which is equal to one if the village is dominated by low castes and zero otherwise.[18] The results are shown in table A2.1 in the appendix. Villages' social structure does not seem to matter for agricultural productivity and the addition of the dummy variable does not change the impact of neighbors' education. It may be due to the fact that I already control for irrigation in the productivity estimation, which is the channel that drives the lower productivity of low castes in high castes dominated villages.

The final argument against explaining the results by systematic differences in

[18]The definition of dominance used here follows Anderson (2011)'s definition: low castes are dominant if they own the largest proportion of land in the village.

terms of wealth and educational returns, comes from the results obtained in table 2.4 which shows that there is no impact of members from the same caste who do not have agriculture as their main activity. Again, given this result, for the positive impact of caste's mean education level to be driven by unobservables, unobservables must be totally uncorrelated within castes. Wealth and educational returns are in fact two determinants of the educational demand which have been shown to be correlated within castes, because of insurance mechanisms (Munshi and Rosenzweig, 2006), and networks effects for accessing jobs (Munshi and Rosenzweig, 2009).

7 Conclusion

This paper takes advantage of the social structure in India to study spillovers of education in farm productivity. The results show that the mean education level of education of members from the same caste is positively related to households' productivity. On the contrary, there is no cross-castes effect. The results also show that within castes, only households who have agriculture as their main activity have a positive impact on households' productivity. This pattern of results helps ruling out that the measured impact is a spurious correlation due to unobservable characteristics. Additionally, this paper shows that spillovers are bigger for wheat than for rice, which is consistent with the fact that learning from neighbors is limited when growing conditions strongly depend on land characteristics.

These findings confirm that education externalities do not only exist in urban contexts and education spillovers do not only occur between workers of the manufacturing and service sectors. There are also spillovers in more traditional sectors such as agriculture. Therefore, improving education in developing countries should continue to be a priority, because education has a multiplicative effect, even in rural areas. Moreover, this paper underlines that education has to be publicly financed, because social returns to education are higher than private returns. The booming of private schools in rural India cannot be a long-term solution.

Maybe even more important, the fact that the external effect of education is happening solely within social groups calls for an equality of access to education across social groups. In India, the law imposing that 25% of classroom seats should be reserved for children from poorer or disadvantaged families in the neighborhood goes into that direction. More generally, affirmative action programs in favor of low castes should continue to be developed.

These findings also have implications for other developing countries, where school enrollment is low and inequality across ethnic groups is also very salient.

Chapter 3

Stigma in affirmative action application? Evidence from quotas in education in India

1 Introduction

Social policies designed to reduce poverty often suffer from low take-up (Currie, 2006). Aizer (2007) for example underlines that half a million children in the United States did not have health insurance whereas they could benefit from Medicaid. Riphahn (2001) finds that 60% of the households eligible for transfers under the German social assistance program did not claim their benefits. Low take-ups can have dramatic consequences in terms of poverty reduction. If people eligible to social programs do not apply, even policies which have proved to be helpful to the poor once they benefit from it will not succeed to alleviate poverty. Understanding the determinants of social programs participation is therefore key to poverty reduction and development in general.

Although not taking the benefits of a social program can appear on a first sight as an irrational behavior, several explanations have been suggested. First, it can come from poor information on the program (Aizer and Currie, 2004). For example, it is possible that people are not aware of the existence of the program, or do not know how to benefit from it. Second, transaction costs may be too high (Bitler et al., 2003; Aizer, 2007). If the procedure to apply is very long or complicated, it is very likely that take-up rates will be low. Finally, eligible people may not apply to avoid the stigma attached to the policy, where stigma is defined as "the disutility arising from the participation in a welfare program per se" (Moffitt, 1983). This disutility comes from a psychological cost which is due to negative self images because of the participation, or to "negative social attitudes towards welfare claimants" (Besley and Coate, 1992). The first two explanations have been well studied empirically (see Currie, 2006, for a review). But empirical evidence on stigma in social programs take-up is scarce, probably because this channel is hard to measure and identify.

This paper aims at filling this gap, by empirically questioning the role of stigma in welfare take-up in India. The social policy that I am studying is affirmative action in higher education. In India, low castes have a privileged access to universities, thanks to quotas that they can apply to. However, they can also choose to not benefit from the quotas, by applying without mentioning their caste. The role of stigma is considered by looking at the impact of households' social status on their probability of applying for reservations in higher education institutions. Under the hypothesis that stigma increases with social status, which is an hypothesis discussed in section 3, finding that households with a high social status apply less to reservations than households with a low social status is consistent with a stigma effect.

Affirmative action in India is a very interesting case study, because it is applied on a very large scale: almost 50% of the seats for students in higher education

institutions are reserved for low castes. However, according to my database,[1] only 15% of the eligible households declare having applied to these quotas in 2005/2006 or 10 years ago. Moreover, the very peculiar social structure of India, the caste system, provides a particularly suitable context to study the impact of stigma on social policies take-up, because castes exogenously determine individuals' social status.

I focus on a specific group which got eligible for reservations quite recently, the Other Backward Classes (OBC).[2] The OBC category is composed of thousands of subcastes or *jatis* whose social status varies from one village to the other depending on secular characteristics. As underlined by anthropologists (Dumont, 1970; Srinivas, 1987), one important determinant of the social status of a jati in a given village is the proportion of land that it owns in this village. Therefore, to measure the impact of social status on the probability of applying for reservations in education, I proxy households' social status by the proportion of land that their jatis own in the village. The distribution of land in villages has the advantage to be historically and exogenously determined as it will be explained in section 3 and consequently to be exogenous to jatis' characteristics. I also exploit the variation of social status across jatis of a same village and across villages for a same jati to control for unobserved factors at the jati and at the village level.

I find that households from jatis who are socially highly ranked in villages are less prone to apply for reservations than jatis with a lower social status. This effect is stronger for richer households. While there can be other explanations than stigma to this effect, I am able to rule out alternative explanations and notably that the effect just captures different returns to education.

This paper is an important contribution to the literature on welfare take-up. By using exogenous factors related to jatis' social position, it allows for a clean identification of the impact of social status, and therefore of stigma. In doing so, it provides empirical support to the theoretical literature on stigma (Moffitt, 1983; Besley and Coate, 1992) and confirms previous descriptive results found in the literature, that stigma prevents individuals from benefiting from social programs (Riphahn, 2001; Stuber and Kronebusch, 2004).

It is also an important input to the literature on affirmative action in India. To my knowledge, it is the first paper to be concerned with the determinants of reservation *application*. Whereas understanding the determinants of take-up is essential to understand the success or failure of a policy, so far the focus has been solely

[1] The ARIS-REDS data.

[2] The OBC is a non homogeneous group constituted of many subcastes or jatis. They are higher in the traditional hierarchy than the Untouchables but they are also considered as economically and socially backward. More information is provided in section 2.2.

on the *impact* of affirmative action programs. Additionally, most of the papers are concerned with reservations in electoral positions (Pande, 2003; Chattopadhyay and Duflo, 2004, see for example). The few papers that focus on affirmative action in education find mixed results. Cassan (2011), using a quasi-natural experiment, finds that affirmative action in education did not have an impact on the education level of scheduled castes. Bertrand et al. (2010), studying labor market outcomes of low castes who benefited from affirmative action to enter universities, find that low castes improve their income by going to the university but less than high castes students.

Finally, this paper brings light into the debate around reservations for OBC. This policy has been accused of favoring the "creamy layers", that is to say the better-offs, who would not have needed reservations to access universities. The results show that stigma partially compensates this effect by preventing households from powerful jatis to apply for reservations.

The rest of the paper is organized as follows. Section 2 describes the Indian context and reservation policies before focusing on the OBC group and social status. Section 3 explains the empirical strategy and discusses its validity. Section 4 provides details on the data and some descriptive statistics. Section 5 shows the results and section 6 discusses the interpretation of the results. Finally section 7 concludes.

2 Contextual background

Affirmative action in India is caste based. This section therefore briefly describes the caste system and its relation to affirmative action before focusing on the OBC and their status in the Indian society in a second part.

2.1 Affirmative action policies in India

The Indian society is divided in a multitude of subcastes or *jatis*.[3] Most jatis are geographically limited and the same jati is rarely present in more than two or three different States, but their scope extends beyond village boundaries. Jatis are hereditary and endogamous groups: one has the same jati as his parents, and has to marry within her/his jati. Moreover, jatis have a hierarchical position which is related to their degree of purity, where the degree of purity notably depends on their traditional occupation. Some jatis with intellectual occupations are very pure and are therefore considered as high castes. Other jatis who have very menial jobs are very

[3]Hereafter I indistinctly use the words subcastes or jatis.

impure and are considered as low castes. However, the hierarchical position is only clear for jatis at the top and jatis at the very bottom. In between, the classification is more flexible and varies locally. The determination of the position of each jati in the traditional hierarchy is further explained in section 3.1.

The degree of purity and the hierarchical position have important consequences in every day life because they determine how members from different jatis interact with each other. Because of their impurity, people from jatis with a low hierarchical position cannot access certain part of villages or cannot share the food of people from higher jatis. The jati's hierarchical position is also an important predictor of individuals' economic status, in part because low castes were restricted to low-skilled occupations but also because they were suffering from discrimination from higher castes, that prevented their economic mobility.

To fight this historically inherited economic and social inequality, affirmative action programs have been put into place in three areas: in elections, in the public sector and in higher education. These affirmative action programs are caste based. One has to be from a jati officially listed to become a beneficiary. Three different categories of jatis have access to affirmative action: the "Scheduled Castes" (SC), the "Scheduled Tribes" (ST) and the "Other Backward Classes" (OBC). The SC, also called Untouchables or dalits are people at the very bottom of the hierarchy. ST refers to tribal people, and OBC is a group which is constituted of hierarchically low jatis, but not as low as the SC.

These three groups did not get access to affirmative action at the same time. Whereas SC and ST have had reservations since Independence, OBC got access to reservations progressively, depending on the State. The first States to implement reservations on the same basis as those for SC/ST were the four southern States (Kerala, Karnataka, Andhra Pradesh and Tamil Nadu) (Galanter, 1978) and the last were the States of the Hindu Belt (Rajasthan, Haryana, Uttar Pradesh, Madhya Pradesh). Since 1993, there has also been reservations for OBC in the Central administration[4] and since 2006 in central higher education institutions.

The reason why OBC got reservations later is because this policy is highly controversial. The OBC constitute almost 50% of the Indian population.[5] Therefore, extending the quotas to this whole population is not anecdotal and lowers the number of remaining seats for the other castes. It is also a very heterogenous population

[4]The word "Central" is used to refer to what is defined at the federal level, and the word "State" to what is defined at the State level.

[5]This estimation comes from the Mandal Commission, which was established in 1979 with the mission of making a report on the status of the OBC. However, this number is controversial, because it has been calculated with figures from the 1931 census, the last census with details by jatis.

in terms of wealth and power. Some OBC had the opportunity to improve their economic status and cannot be considered anymore as disadvantaged. Moreover, among the OBC, some jatis as a whole enjoy locally very influential positions due to their landholding position.

2.2 The application procedure to reservations

To benefit from affirmative action in higher education, individuals have to go through the normal application procedure, but they are only in competition with individuals from their own caste group. The grades that individuals have secured at the high school examination first determine their ability to apply for universities, because universities have thresholds under which they do not accept applications. The threshold for low castes is often lower. Once having applied, candidates have to take a competitive exam. All the candidates are classified according to what they achieved at this test, but separate lists are made for SC, ST and OBC that declared to be so.

This is at the beginning of the process, when applying to universities, that individuals have to specify if they want to be considered under the SC, ST, OBC or general category. While proofs of caste status are required when applying as an SC, ST or OBC, anyone can apply under the general quota. However, if eligible, it is in the interest of individuals to compete under the quotas, because the cut-off is always lower in those categories. Bertrand et al. (2010), studying admission into engineering schools, find for example that the cut-off score for the SC was 182, 419 for the OBC and 490 for the general category.

2.3 OBC, stigma and social status

As for other social programs, it has been observed that those who benefit from reservations suffer from social stigma. Gudavarthy (2012), who studies the stigmatization of reservations, reports that Scheduled Castes are often referred as *sarkar ke damad* (sons in law of the government). They are also accused to be less competent and the students who got access to universities through reservations are ostracized by high caste students. For OBC, stigma is also related to the fact that reservations have been historically designed for the Scheduled Castes. While OBC are also low castes, they do not suffer from untouchability and being compared to the Scheduled Castes can be seen as degrading.

In this context, stigma may affect the probability of applying for reservations in higher education in three ways. The impact can be very direct. For individuals enjoying a relatively high status in their village, living their good situation in terms

of status for a situation where they will be badly considered by the others is costly. The way stigma affects application to reservations can also be more indirect. In India, social status has a value in itself (Bloch et al., 2004) and behaviors meant to improve or signal social status have frequently been observed in diverse situations. In particular, some low castes tend to adopt the behavior and customs of high castes in order to improve their social status. This phenomenon has been called "Sanskritization" and one of the most reported example is the adoption of high castes' eating habits (Srinivas, 1956). But behaviors aiming at signaling social status have also been described in other contexts, such as in wedding expenditures (Bloch et al., 2004) or in the time allocation of women (Eswaran et al., 2013). Therefore, adopting a "low caste behavior" such as applying for reservations does not go in the right direction for subcastes trying to improve their status. Finally, stigma may also be private. Individuals may not want to benefit from reservations because they do not value as much their accomplishment if it is thanks to quotas.

Given the ways stigma may affect reservation application, my analysis of the impact of social status focuses on OBC. SC and ST are so low in the traditional hierarchy that we do not expect stigma to prevent them from applying for reservations. On the contrary, OBC, because of their intermediary position in the traditional hierarchy have more room for status improvement. And in fact, according to the literature, OBC attach more importance to status than other groups. Khamis et al. (2012) for example find that OBC spend 8 percent more on visible consumption than high caste groups, after controlling for income.

For the same reasons, in this paper the hypothesis is that the cost of adopting a stigmatized behavior like applying to reservations increases with social status. Among OBC, individuals from subcastes with a low social status have little to loose in terms of reputation, whereas people with a high social status suffer much more from a loss in how others consider them. Under this hypothesis, we expect that other things being equal, if there is stigma, people with a higher social status will be less prone to apply for reservations.

The next section explains how social status is measured.

3 Identification strategy

To measure stigma, I look at the impact of social status on application to affirmative action in higher education. This section explains my measure of social status before turning to the empirical specification *per se* and the identifying assumptions.

3.1 How to measure social status?

3.1.1 Land as a determinant of social status

As explained in section 2, jatis have a hierarchical position which determines the social status of their members. The hierarchy is clear at the broad level: high castes have a higher social status than OBC who have a higher social status than SC and ST. However, the hierarchical position of a given jati within these broad categories is not as clear. In fact, the relative position of a specific jati in comparison to another jati can vary locally.

Anthropologists have underlined that this variation of status within broad groups is largely governed by patterns of land ownership in villages. Srinivas (1987) for example explains that the life of Indian villages is governed by the jati which is economically the most powerful in the village, in other words the jati which owns the most land. This subcaste is called the "dominant caste". Dumont (1970), also underlines the importance of land control for dominance.

Following this anthropological literature, I use jatis' land ownership as a measure of jatis' local social status. More precisely, households' social status is measured by the *proportion of land owned by their jati in the village*. Two things are important to note. First, it is the *proportion* and not the *mean* land owned which is used, because jatis do not get power in a village through how much each member in the village possesses, but by how much the jati as an entity possesses.[6] Second, measuring social status through the proportion of land owned by the jati is a *proxy*. They are other determinants of social status than land in villages. However, in section 6 I provide further empirical evidence that the proportion of land owned by the jati is a good measure of social status. Jati's land ownership has also already been used in the economic literature to measure power relations in villages (Anderson, 2011).

3.1.2 Exogeneity of land settlement

In addition to really reflecting power relations in the village, for the proportion of land to be a good proxy of jatis' social status, it also needs to be exogenous to jatis' characteristics. If some jatis have more land than others because they have different characteristics, land may also proxy for these characteristics. Two main evidence support the exogeneity of land ownership pattern.

[6]The hierarchical position of a jati is therefore also correlated to how numerous the number of jati members in the village, as underlined by Dumont (1970). However, the proportion of land is preferred over the size of the jati to measure social status, because it is a closer proxy for extreme situations. For example, a very numerous jati without land will have a low social status, whereas members from a jati which has few representatives in the village but possesses the majority of land in the village will enjoy a relatively high social status.

First, land ownership in villages is historically determined and has barely changed since the land reforms which took place after Independence. This fact has been documented at the State level by Besley and Burgess (2000) and at the village level by Anderson (2011). Anderson (2011)'s work focuses on Northern India, but the ARIS-REDS data that I am using here show a similar pattern for Southern States. Households' heads in the 2006 round had been asked about land transactions in their household since they became heads. In total, less than 1.6 % of the households declare having sold or gifted some land during this period. Furthermore, among those who did, 33% transferred the land to family members of friends. Given that most of the relationships are intra-jati, we can globally interpret this as a transfer to persons from the same jati. Therefore, if landownership has been modified, it is only marginally, given that transfers are few in quantity and are mainly to other members of the same jati.

Second, land ownership patterns could also have been modified because of migration. If migration is high and does not affect jatis evenly, it can create endogeneity issues. But migration in rural India is very low. According to the Indian census of 2001, only 4.7% of the total population of India was born outside of the State of residence. This point has been notably studied by Munshi and Rosenzweig (2009) who argue that this is due to strong insurance mechanisms in jatis. Moreover, 43.8% of total migration is migration for marriage purposes.[7] Women migrate to live in their husbands' family and do not inherit from the land. Consequently, migration did not dramatically affect land distribution across jatis in villages.

3.1.3 Variation of land ownership pattern

Finally, using the proportion of land as a measure of social status requires that there is variation in the land ownership pattern across jatis and across villages. To show that this is the case, table 3.1 takes the example of the jati composition of two villages situated in the State of Chhatisgarh. Village 1 has seven OBC jatis and village 2 has six. In both villages, there is variation of land ownership between OBC jatis of the same village: in village 1, two jatis own respectively 21% of the land and 47% of the land while the five remaining jatis own each less than 5%. In village 2, the situation is a little more even between the jatis: one jati owns 37% of the land in the village and the other jatis own between 0 and 12% of the land. There is also variation for a same jati across villages: while in village 1 the Kewat own 21% of the land, in village 2 they only have 6% and are among the jatis who have the least land. On the contrary, the Teli who have only 1% of the land in village 1 have 37% of the land in village 2.

[7] Census of India 2001.

Table 3.1: OBC jati composition

Jati	Proportion of land owned in %	
	Village 1	Village 2
Ahir	0.02	0.12
Kallar	0.03	
Kewat	**0.21**	**0.06**
Kurmi	**0.47**	**0.08**
Lohar	0.05	0.01
Nai	0.01	0.02
Parit	0.03	
Patwa		0
Teli	**0.01**	**0.37**

The two villages are in the State of Chhatisgarh. The total amount of land is not equal to 100% because only the OBC jatis are reported.

The land distribution of these two villages is representative of the situation in the other villages of my sample. This is this variation in land ownership that I exploit to identify the impact of social status on the probability of applying for reservations, as I explain in the next section.

3.2 Empirical specification

The role of stigma is studied by estimating the impact of jatis' social status on households' probability of applying for reservations, where jatis' social status is measured by the proportion of land owned by jatis in villages. The probability of applying for reservations is measured at the household level, because the data provide the information for this level of aggregation. The empirical specification is as follows:

$$y_{ijv} = \alpha + \beta LAND_{jv} + \gamma X_{ijv} + \sigma_v + \theta_j + u_{ijv} \tag{3.1}$$

where y is equal to one if a member (including the head) of household i of jati j in village v applied for reservation and zero otherwise. $LAND_{jv}$ is the proportion of land owned by the jati in the village and X_{ijv} is a vector of households characteristics.

Unobservables are controlled for at two levels. First, I use the variation of status *across OBC jatis* in a same village to control for village level unobservables, by adding villages dummies σ_v. This accounts for characteristics which are common to all members of the same village, such as access to primary and secondary education, job opportunities or land fractionalization. Second, if the status of a given jati is due to its unobservable characteristics, the measured effect of social status on

reservation application is most likely to be biased. To account for that I exploit the fact that there is also exogenous variation of status in a same jati *across villages*, and I control for jatis unobservables represented by θ_j in some specifications.

The equation is estimated with OLS. Although it poses restrictions on the functional form of the relation between social status and application for reservations, with this estimator village and jati dummies can be added at the same time. Robustness checks are provided using conditional logit when jati dummies are not included, but the number of observations per jati is too small to include jati dummies with a logit estimator. The error term u_{ijv} is clustered at the jati in the village level. Given that the proportion of land owned by the jati measures its local social status, if there is stigma we expect that the probability of applying for reservations decreases with the proportion of land owned by the jati in the village

4 Data and descriptive statistics

4.1 Data

The data used to conduct this study are from the 2006 round of the ARIS-REDS database from the National Council of Applied Economic Research (NCAER). Since 1971, the NCAER has been conducting household surveys along with village surveys in 259 villages in the 17 major States of India.

The 2006 round is a very peculiar one. Along with the usual questions asked to a sample of households, a complete census of all the households in every village has been conducted. This is this census that I am using here, because it provides a complete picture of villages social stratification. In particular, households were asked the name of their jati, so I am able to construct the subcaste group of each household.

Households were also asked about their application to reservations in the current year of the survey and 10 years ago. I use this information to construct my dependent variable: it is a dummy variable equal to one if a member of the household applied to reservations in education in 2006 and/or 10 years ago and 0 otherwise.

The variable of interest is the proportion of land owned by the jati in the village. To compute this variable I aggregate by jati the data on households' land ownership ten years ago. Finally, the census also provides households' demographic characteristics, that I am using as control variables. In particular, I control for the age of the household's head, his/her gender, his/her education level, the number of members in the household, and the amount of land owned by the household in log.[8]

[8]The logarithm transformation is used to take into account the fact that the land variable is very skewed to the left. As several households do not have land this variable is transformed after

4.2 Sample

My dependent variable relies on information on reservation application ten years ago and in 2006. I therefore concentrate on the States who had reservations for OBC more than 10 years ago. In particular, I exclude Uttar Pradesh, West Bengal, Orissa, Madhya Pradesh, Rajasthan and Haryana because they adopted OBC reservations too recently, as explained in section 2. I also exclude Punjab, because the policy in favor of the OBC is very marginal: the quotas for OBC is only of 5 %, and Himachal Pradesh because there are not enough observations.[9] My study consequently focuses on the 9 States represented in figure A3.1 in the appendix. They are almost all in South India.

The identification strategy also requires that there is enough variation in each village in the variable of interest, the proportion of land owned by the jati. I therefore focus on villages where there are at least 3 OBC jatis. The final sample is composed of 28998 OBC households distributed in 92 villages. When jati dummies are included, I also need variation in the variable of interest within jatis. In those specifications, I only include jatis which are present in at least three villages.

4.3 Descriptive Statistics

Table 4.2 shows summary statistics separately for the OBC households who applied to reservation and for those who did not.

Table 3.2: Descriptive Statistics per application status

	No Application				Application				Diff	Sign
	Mean	SD	Min	Max	Mean	SD	Min	Max		
Age	46.12	13.6	19	106	47.27	12.71	19	105	-1.14	***
Female	0.11	0.31	0	1	0.10	0.29	0	1	0.01	**
Education	4.44	4.08	0	19	6.67	4.42	0	19	-2.23	***
Size of the HH	4.86	2.55	0	63	4.92	2.58	0	41	-0.61	N.S.
HH Land owned	1.09	13.70	0	1950	1.45	9.86	0	650	-0.36	*
Jati prop land	0.28	0.28	0	0.87	0.29	.29	0	0.87	0.01	*
N	24085				4913					

Land owned by the household is in acres.

As we can see, there is selection in application. Households who apply for reservation are more educated (in mean the head of the household has 2.2 more years of education), and they are richer. Given that reservations are supposed to help the most disadvantaged households, this result can seem surprising. However it is in line with the literature (Bertrand et al., 2010) and with the debate on the "creamy

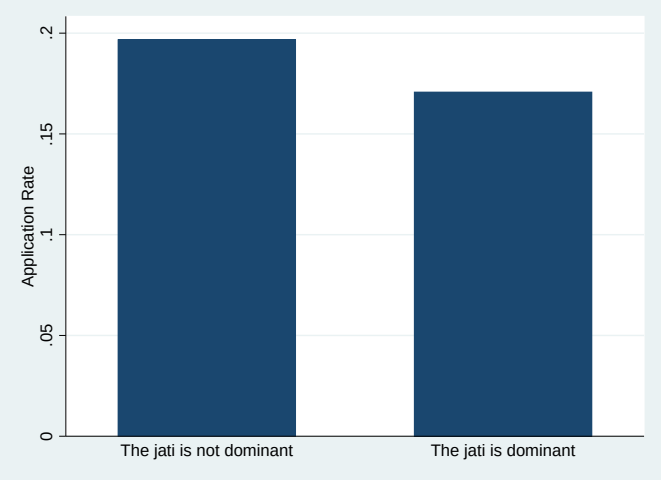
having added one to the actual value.

[9]There are only 5 OBC households in Himachal Pradesh in the data.

layers" explained in the introduction. The proportion of land owned by the jati in the village does not seem to differ strongly depending on the application status.

But if we only look at jatis with a high status in at least one village, the picture is different. Figure 3.1 shows the application rate to reservations in education among jatis which own the largest proportion of land (or *dominant jatis*) in at least one village.

Figure 3.1: Application rate of jatis which are dominant in at least one village



Note: this graph is constructed with OBC jatis that are dominant in at least one village and that are represented in at least in 3 villages. The number of households in this sample is 14027, where 4819 are in villages where the jati is not dominant, and 9208 where the jati is dominant. The difference between the application rate of the two groups is statistically significant from 0 at a 1% level.

The left bar represents the application rate of dominant jatis in villages where they are not dominant and the right one in villages where they are dominant. When the jati is dominant in the village, the application rate of households belonging to this jati is lower than when the same jati is not dominant. So it seems that there is a difference of behavior among people in the same jati, depending on the social position of the jati in the village.

5 Results: the determinants of reservation application

5.1 Main results

Table 3.3 shows the results of the estimation of equation 3.1. In columns (1) and (2) the equation is only estimated with controls at the household level. Column (3) to (6) show the results with the proportion of land owned by the jati.

Table 3.3: Main results: households characteristics and stigma in reservation application

Dependant Variable:	**Application status to reservation in education**					
	(1)	(2)	(3)	(4)	(5)	(6)
Estimator:	OLS	OLS	OLS	OLS	OLS	Logit
Age (Years)	0.000729***	0.000924***	0.000748***	0.000936***	0.000930***	0.00737
	(0.000250)	(0.000283)	(0.000250)	(0.000283)	(0.000285)	(0.0139)
Female	0.00536	0.00494	0.00536	0.00512	0.00490	0.0696
	(0.00655)	(0.00719)	(0.00658)	(0.00721)	(0.00722)	(0.132)
Number of years of	0.0107***	0.0114***	0.0107***	0.0114***	0.0114***	0.107***
schooling & college	(0.00186)	(0.00225)	(0.00186)	(0.00226)	(0.00228)	(0.0309)
Household Size	0.00658***	0.00685***	0.00657***	0.00686***	0.00685***	0.0683**
	(0.00161)	(0.00186)	(0.00160)	(0.00186)	(0.00186)	(0.0348)
HH land owned (in	0.00722	0.00807	0.00930**	0.00887*	0.00880*	0.0575
log)	(0.00453)	(0.00504)	(0.00449)	(0.00507)	(0.00507)	(0.0647)
Jati prop land owned			-0.0455**	-0.0544**	-0.0555**	-0.442*
			(0.0188)	(0.0256)	(0.0258)	(0.261)
Jati mean education					0.00190	
					(0.00397)	
Village dummies	yes	yes	yes	yes	yes	yes
Jati dummies	no	yes	no	yes	yes	no
Jati control variables	no	no	no	no	yes	no
N	28998	23102	28998	23102	23102	26286
r2	0.319	0.338	0.320	0.338	0.338	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are corrected for clustering at the jati in the village level. Age, female and education are those of the household head. The jati level variables are the mean values of age, female, education and household size calculated over households from the same jati in the village. The concerned household is excluded from the mean calculation.

The results on the households characteristics are as expected. Column (1) shows the estimation with village dummies and column (2) shows the estimation with village and jati dummies. The probability of applying for reservations significantly increases with the age of the household's head, the size of the household, the number of years of schooling of the household's head and the amount of land owned by the household. All the results are robust to the inclusion of jati dummies and the coefficients are stable, so jati unobservables do not seem to create endogeneity problems. Having a female at the head of the household on the contrary does not matter. Whereas the age of the household's head and the household size mechanically increase the probability of applying for reservations because the dependent variable is at the household level,[10] the results on the education of the household's head and on

[10]The age of the household's head has a positive impact probably because it is positively correlated with the probability that the household has members in age of applying for reservations.

land owned by the household confirm what was shown in the descriptive statistics. Households who apply for reservations are significantly more educated and are richer than those who do not apply.

Column (3) to (6) study social effects in reservation application by looking at the impact of the proportion of land owned by the jati on the probability of applying for reservations. Column (3) and (4) show the basic estimation of equation (1), without and with jati dummies. In both specifications, the coefficient on the proportion of land owned by the jati in the village is negative and significantly different from zero at a 5% level. In column (3), the estimation is made only with village dummies, showing that the relative social position of the jati within the village has a strong impact on the probability of applying for reservations. The higher is the jati in the village hierarchy, the less likely are households from this jati to apply. Column (4) shows the results with both village dummies and jati dummies. The coefficient is slightly bigger in absolute value. It confirms that the measured impact of the jati level variable in column (3) is not driven by jati unobservables. Households from the same jati, but in different villages do not have the same probability of applying for reservations: *ceteris paribus*, households whose jati has more land in one village have a lower probability of applying for reservations in this village than in a village where this jati owns less land. Column (5) includes the control variables at the jati in the village level, to see if the proportion of land owned by the jati does not capture the effect of other jati's characteristics. I only report the coefficient for jatis' mean education level, but none of the control variables at the jati level is significant. The coefficient on the proportion of land owned by the jati is still negative and significantly different from zero. Finally column (6) tests the robustness of the results by using a conditional logit to estimate the impact. The result is consistent with what was previously obtained.

The results obtained in column (3) to (6) confirm that the proportion of land owned by the jati in the village has a negative impact on the probability of households from that jati to apply for reservations. As the proportion of land owned by the jati measures social status, under the hypothesis that stigma increases with social status this result is consistent with a stigma effect. However, it is also possible that the demand for education changes with the proportion of land owned by the jati. If this is the case, the same results can be obtained without any stigma effect. This alternative explanation is considered in the next section.

5.2 Alternative explanation: different returns to education

The results previously obtained are consistent with a stigma effect, but there is another valid alternative explanation: it can also be that people who are from a jati which possesses a high proportion of land in the village have lower returns to education. In this case, people from jatis with a high proportion of land apply less for reservations because they do not want to go to the university, and not because of stigma.

This alternative explanation of different returns to education is partially addressed by the fact that the amount of land owned by the household is already controlled for. If differences in land ownership were leading to different returns in education, then controlling for households' land should have captured the impact of the proportion of land owned by the jati. However, it can still be that the proportion of land owned by the jati captures outside opportunities. For example, households from jatis with a lot of land do not need to go to the university to get a job, because even poor households will be able to work on the land of other households from the jati. This section aims at ruling out this alternative explanation, using several robustness checks. The results are shown in table 3.4. I only report the estimations with village and jatis dummies, but they are similar to the ones with only village dummies.

Table 3.4: Robustness checks

Dependant Variable:	**Application status to reservation in education**			**HH's head educ**
	(1)	(2)	(3)	(4)
Group considered:	OBC	OBC	SC/ST	OBC
HH land owned (in log)	0.00641	0.00924	0.0314**	0.738***
	(0.00592)	(0.00646)	(0.0123)	(0.0844)
Jati prop land owned	-0.0517**	-0.0633*	0.119	0.325
	(0.0257)	(0.0363)	(0.0950)	(0.277)
Village dummies	yes	yes	yes	yes
Jati dummies	yes	yes	yes	yes
HH control variables	yes	yes	yes	yes
Jati control variables	no	no	no	no
Occupation	yes	no	no	no
N	23080	12213	5274	23102
r2	0.346	0.358	0.382	0.284

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The regressions are estimated with OLS and standard errors are corrected for clustering at the jati in the village level.

In column (1), I control for the primary occupation of the household's head. Given that the returns to education are highly correlated to the occupation that the household has, controlling for this variable should lower the impact of the jati's

proportion of land owned if it captures different returns to education. It is not the case, adding the occupation does not change the impact of the proportion of land owned by the jati.

Column (2) checks if focusing on households whose members have a high probability of getting education changes the results. The probability of getting education is highly correlated to the education level of one's parents. So in column (2) I restrict the sample to households whose members have a high probability to get education, that is to say households where the household head is educated. There are not enough households where the household's head has a university degree so I focus on households whose head has at least finished primary school. In this restricted sample, the impact of the proportion of land owned by the jati is less precisely estimated, but still significant at a 10 % level and the coefficient is broadly the same as in the main estimation.

In column (3) I look at the impact of the proportion of land owned by the jati on the Scheduled Castes and Scheduled Tribes' (SC/ST) probability of applying for reservations. As explained in section 2, SC/ST have a very low social status, so the stigma attached to reservations should not affect their application. If the negative impact of the proportion of land owned by the jati is due to stigma, then when considering SC/ST the impact should not be negative anymore. But if the proportion of land owned by the jati actually captures different returns to education, then the proportion of land owned by the jati should be negative for both OBC and SC/ST. The result in column (3) contradicts the interpretation of previous findings as driven by different returns to education: for SC/ST the impact of the proportion of land owned by the jati is positive (but not significant).

Column (4) looks at the correlation between the household's head level of education and the proportion of land owned by the jati in the village. If returns to education diminish with the proportion of land owned by the jati, then we would expect that households' head from jatis with a high proportion of land have a lower level of education.[11] The results from column (4) contradict this hypothesis: the level of education of households' heads is not correlated with the proportion of land owned by the jati.

All these robustness checks contradict the interpretation of the main results as a lower return to education for households from jatis with a high proportion of land in the village.

[11] Under the hypothesis the gap between returns in agriculture and in other sectors has not changed over time.

5.3 Heterogeneity of the impact

I now look at the heterogeneity of the impact. Is the impact of social status uniform among households or does it differ depending on households demographic characteristics? I consider two sources of heterogeneity. First, the impact of social status can depend on the education level of the household's head. One can think for example that more educated people care less about their social status. Second, the impact of social status can depend on the wealth of the household itself. To consider these two questions, I interact the proportion of land owned by the jati with the education level of the household's head on one hand, and with the amount of land owned by the household (in log) on the other hand. Table 3.5 shows the results of including these interaction terms in the main specification.

Table 3.5: Heterogeneity of stigma

Dependant Variable:	**Application status to reservation in education**			
	(1)	(2)	(3)	(4)
Number of years of schooling & college	0.0135***	0.0138***	0.0108***	0.0115***
	(0.00242)	(0.00300)	(0.00186)	(0.00226)
Educ * prop land jati	-0.0107**	-0.00893		
	(0.00499)	(0.00601)		
Jati prop land owned	0.0125	-0.00486	-0.0195	-0.0363
	(0.0316)	(0.0389)	(0.0183)	(0.0260)
HH Land owned (in log)	0.0101**	0.00967*	0.0321***	0.0269***
	(0.00448)	(0.00501)	(0.00798)	(0.00740)
HH Land owned * prop land jati			-0.0631***	-0.0507***
			(0.0174)	(0.0167)
Village Dummies	yes	yes	yes	yes
Jati Dummies	no	yes	no	yes
HH control variables	yes	yes	yes	yes
Jati control variables	no	no	no	no
N	28998	23102	28998	23102
r2	0.32	0.34	0.32	0.34
Interact & jati var	0.008	0.051	0.000	0.002
Interact & HH var	0.000	0.000	0.000	0.001

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The regressions are estimated with OLS and standard errors are corrected for clustering at the jati in the village level. The two last lines of the table report the p-value for the joint significance of the interaction term and the variable with which jatis' proportion of land owned is interacted.

Column (1) and column (2) show the results when the education level of the household's head is interacted with the proportion of land owned by the jati in the village. Column (1) shows the results without jati dummies and column (2) with

jati dummies. The interaction term is significant in the first specification but is not robust to the inclusion of jati dummies. The impact of social status does not seem to differ across different levels of education.

On the contrary, the results on the interaction of the proportion of land owned by the jati with the area of land owned by the household are very interesting. In both specifications with and without jati dummies, the interaction term is negative and strongly significant. This result shows that the *negative* impact of the proportion of land owned by the jati *increases* (in absolute value) with the amount of land owned by the household. More clearly, this result shows that the impact of social status is stronger for richer households than for poorer households. It also means that the *positive* impact of the amount of land owned by the household *decreases* with the proportion of land owned by the jati. The impact even becomes negative when the household is from a jati which owns a large proportion of land in the village. In other words, while rich households of not so powerful jatis are more prone to apply for reservations than poor ones, rich households from powerful jatis are less prone to apply than poor households. Therefore, it seems that stigma prevents rich households with high status from applying for reservations.

This result is an additional argument against the idea that the negative impact of the proportion of land owned by the jati is driven by different returns to education across jatis and not by stigma. Indeed, if the proportion of land captures a lower demand for education, then we would expect that rich households in powerful jatis are less affected than poor ones. The rationale is that rich households should not be impacted by the proportion of land owned by their jati if it captures outside opportunities because they already have work opportunities in their own household. But rich households (those with more land) are actually more affected by the proportion of land owned by their jati.

6 Discussion

Section 4 underlines strong effects for OBC households of the proportion of land owned by their jati on affirmative action application. These results are robust and do not seem to be driven by village or jati level unobservables. In this section I continue the discussion began in section 5.2 on the interpretation of the results. In the first subsection I provide empirical evidence that the proportion of land owned by the jati really measures jatis' social status. In the second subsection I discuss other explanations than stigma to the impact.

Table 3.6: Political outcomes

Dependant Variable:	**Candidate to a local election**			
	(1)	(2)	(3)	(4)
HH land owned (in log)	0.0114***	0.0140***	0.0114***	0.0139***
	(0.00283)	(0.00347)	(0.00282)	(0.00347)
Jati prop land owned	0.0151***	0.0158**	0.0183***	0.0219**
	(0.00538)	(0.00785)	(0.00627)	(0.00863)
Jati population			-0.00000772*	-0.0000168*
			(0.00000456)	(0.00000948)
Village Dummies	yes	yes	yes	yes
Jati Dummies	no	yes	no	yes
HH control variables	yes	yes	yes	yes
Jati control variables	no	no	no	no
N	28998	23102	28998	23102
r2	0.0451	0.0615	0.0452	0.0615

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The regressions are estimated with OLS. Standard errors are corrected for clustering at the jati in the village level.

6.1 Ownership of land and political outcomes

The measure of social status through jatis' land ownership comes from the anthropological literature as explained in subsection 3.1, and has been previously used in the economic literature (Anderson, 2011). However, one can still be doubtful about its relevance in contemporary India. To give empirical credit to this measure, I study the relation between the proportion of land owned by the jati and another outcome to which social status is also strongly related, being a candidate to elections.[12] Therefore, if the proportion of land owned by a jati actually measures its members' social position in the local hierarchy, we expect people from this jati to have a higher probability to be candidates[13] in local elections other things being equal.

The equation estimated is the same as before, except that the dependent variable is now $candidate_{ijv}$:

$$candidate_{ijv} = \alpha + \beta LAND_{jv} + \gamma X_{ijv} + \theta_j + \sigma_v + u_{ijv} \tag{3.2}$$

where $candidate_{ijv}$ is equal to one if someone of the household has been a candidate to any local election (panchayat and pradhan) and zero otherwise.

In column 1, the equation is estimated with village dummies and in column 2 with village and jati dummies. In both specifications, the amount of land owned by

[12] In India, it has been shown that when elections are not reserved to a specific caste group, local electoral positions are often taken by members from subcastes who socially dominate village life (Witsoe, 2009; Anderson et al., 2011).

[13] Household level informations on election output are not available.

the jati is positive, showing that being from a jati which owns a lot of land in the village increases the probability of being a candidate.

However, given that the proportion of land owned by the jati is also positively correlated to how numerous the jati is, it is not surprising that being from jatis with a high proportion of land increases the probability of being a candidate. To check if this is the size of the jati that is driving the impact, I run the same regressions but with the population of the jati in the village as an additional control variable. The results in column (3) and (4) confirm what was obtained before, the probability of being a candidate is significantly related to the proportion of land owned by the jati in the village.

These results underline that the proportion of land owned by the jati seems to be a good measure of social status.

6.2 Additional alternative explanations

I shall now discuss the interpretation of the negative impact of the proportion of land owned by the jati as a stigma effect. One alternative explanation, the different returns to education across jatis, has already been considered in section 5.2. Here I look at two alternative explanations besides stigma. First, the proportion of land owned by the jati could capture different opportunity costs across jatis. Second, the effect may capture other obstacles to reservation application than stigma.

6.2.1 The opportunity cost explanation

In this paper, I interpret the measured impact of the proportion of land owned by the jati as a stigma effect. The fact that people from jatis with a big proportion of land may have different returns to education than other jatis could also explain the impact, but it does not seem to fit the data as shown in section 5.2. Another related alternative explanation could be that the situation of people whose jati has a lot of land is different, not because of the land directly, but because of their higher social status. It can be that they have access to more resources, or are better connected, and consequently have a higher opportunity cost to apply for reservations, because they have a higher opportunity cost to leave their village. In particular, Anderson (2011) shows that low castes in villages where the majority of land is owned by low castes have a higher agricultural productivity than in villages dominated by high castes. She underlines that it is due to a better access to irrigation for low castes in low castes dominated villages. In high castes dominated villages, high castes prevent low castes from accessing water sources.

Here, it could consequently be that low castes with a dominant position in the

Table 3.7: The opportunity cost alternative explanation

Dependant variable:	Application status to reservation in education		Agricultural production	
	(1)	(2)	(3)	(4)
HH land owned (in log)	0.00878*	0.0129**	0.0367	-0.00456
	(0.00518)	(0.00598)	(0.0281)	(0.0296)
Jati prop land owned	-0.0455**	-0.0538**	-0.0868	0.141
	(0.0188)	(0.0256)	(0.0936)	(0.233)
Village Dummies	yes	yes	yes	yes
Jati Dummies	no	yes	no	yes
HH control variables	yes	yes	yes	yes
Jati control variables	no	no	no	no
Irrigation dummy	yes	yes	yes	yes
Agr control variables	no	no	yes	yes
N	28982	23086	970	723
r2	0.320	0.338	0.883	0.903

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The regressions are estimated with OLS. Standard errors are corrected for clustering at the jati in the village level. The agricultural control variables include the area of land cultivated (in log), the number of days worked on the field, if the household has mechanized or non-mechanized equiment, the proportion of land that is owned and the proportion of land that is irrigated.

village have a better access to irrigation, or other agricultural inputs than the other low castes in the village. If this is the case, they have a higher productivity and a higher opportunity cost to leave the village. I test this explanation in table 3.7. In column 1 and 2, I look at if adding a dummy for wether their land is irrigated or not changes the impact of the proportion of land owned by the jati. It is not the case, the coefficients are similar to those in table 3.3. In column 3 and 4, I test if households in jatis with a high proportion of land have a higher agricultural productivity. As the agricultural production is only available for a subsample of households, the number of observations is much smaller. However, the proportion of land owned by the jati does not seem to matter for agricultural production. Consequently, a different opportunity cost across subcastes (at least regarding agriculture) does not seem to be driving the results.

6.2.2 The information channel

The theoretical literature underlines other possible social effects than stigma in welfare take-up (Moffitt, 1983). In particular, information has been shown to be an important determinant of program participation (Aizer, 2007; Heckman and Smith, 2004). Information sharing is consequently another channel through which the group could affect reservation application. One could argue that the measured social effect actually captures this information channel. However, several facts contradict this

argument.

First, we would expect households from more powerful jatis -jatis with a higher proportion of land- to be better informed. So if the proportion of land owned by the jati captures information effects, we would expect that being in a jati which owns a high proportion of land would have a *positive* impact on the probability of applying for reservations.

Second, it has been shown that in India, higher educated people have better access to information (Foster and Rosenzweig, 1996). Therefore, if there are social effects related to information sharing, we expect that the mean education level of neighbors has an impact on the probability of applying for reservations, or that the impact of social status diminishes with the education level of the household's head. In column (5) of table 3.3 jatis' mean education level is not significant and its inclusion does not affect the coefficient of the proportion of land owned by the jati. Moreover, columns (1) and (2) of table 3.5 does not show any decrease of the impact with the households' head education level. Consequently, the results that I obtain do not seem to be driven by an information effect.

7 Conclusion

Determining who among the ones who have the right to benefit from welfare programs actually take advantage of them is very important for policy designs. In this paper, I am concerned with a very controversial program, an affirmative action program in favor of disadvantaged castes in India in higher education. This program is addressed to the "Other Backward Classes", an heterogeneous entity composed of groups with very different social and economic status. Given that the jatis (subcaste groups) concerned by this program are hierarchically higher in the traditional ranking of castes than the groups historically targeted for affirmative action (the untouchables), stigma may actually have a role in the choice of applying to reservations.

To study that, I look at the impact of households social status on their probability of applying for reservations. I find that being from a locally powerful subcaste has a negative impact on reservation application. I also find that this impact is stronger for households with more land. These results are not driven by village or jati unobservables and do not capture different returns to education across jatis. It therefore seems that stigma may be at stake in keeping households from locally high subcastes to apply for reservation.

This paper consequently shows that part of the low-take up that we observe in welfare programs may be due to stigma. But although stigma may have dramatic

effects when it affects the take-up of health-related programs, in this specific context the consequences do not seem as tragic. Indeed, the take-up of the lowest castes, the SC/ST, is not affected. Moreover, among the OBC, it only affects the richest households from the most powerful jatis. In quotas in higher education, stigma therefore seems to play the role of control barriers to prevent the better-offs from applying to programs that are not designed for them. However, it is important to keep in mind that even if stigma only reduces the take-up rate of the highest castes, stigma may still affect the well-being of those who choose to benefit from it, and as such needs to be taken into account by policy makers. Further research is needed to clarify this issue.

Chapter 4

How to get a job in the public sector? The role of local politics and caste networks in affirmative action programs in India

1 Introduction

Past discrimination against certain people on the basis of their ethnicity, religion or gender has had an important impact on current inequalities. Because those groups were restricted to certain occupations or deprived from certain rights, today, they are lagging behind their country counterparts in terms of socio-economic outcomes. This is for example the case for the African Americans in the US, the Malay in Malaysia, the indigenous population in Brazil or women all over the world. To make up for this situation, a lot of countries have set up affirmative action programs in favor of those groups.

In India, where certain individuals were discriminated against because of their caste, affirmative action programs are implemented on a very large scale. The extent of affirmative action is particularly spectacular in public employment where since Independence, depending on the State, between 22.5% and 49.5% of the jobs in the public sector have been reserved for low castes.[1] However, each year a non-marginal proportion of these quotas remains unfilled and low castes are still underrepresented in the public sector. Two main reasons can explain this failure in filling the quotas. The first reason is that although there are candidates, they are not considered as "suitable" (Jaffrelot, 2011). It can be sometimes true, but it is often used as an excuse by recruiters that do not want to hire low castes, because after several unsuccessful recruitments the position is opened up to all castes. The second reason, which is the one in which this paper is interested in, is that it is not easy to access these jobs, because the recruitment is highly discretionary. Chandra (2004) reports that except for high skilled positions which are filled by competitive exams, a high proportion of the employees at low skilled positions, which constitute almost 95% of the jobs in the public sector, are recruited directly by the offices concerned. This discretion in hiring makes it difficult for low castes individuals without connections to access these jobs and therefore benefit from reservations in the public sector. Chandra (2004) describes the case of an untouchable in a village in Uttar Pradesh, who is eligible for quotas in public sector employment but does not even try to benefit from them because "had he tried to escape his circumstances by securing regular employment in the public sector, he would have needed "contacts"". This difficulty in getting benefits from the State without connections not only concerns reservations in public employment, but a large range of public programs. Therefore, intermediaries are commonly used to mediate people's access to State institutions (Witsoe, 2012). Among these intermediaries, political leaders, who have power over administrative decisions, play an important role.

[1] Section 2 provides an overview of the organization of the social system in India.

The question I am concerned with in this paper is if households that are better connected to the State through their caste networks have an easier access to reserved jobs in the public sector. Because I do not know if households got the job after applying, I study the impact of caste networks on *application* for reserved jobs. More concretely, I look at the impact of having someone from the same caste as a local elected leader on the probability of applying for reservations in the public sector. What is the mechanism? Elected leaders in India do not concretely have the power to attribute administrative jobs. However, they have a certain control over bureaucrats, because they can transfer them across different posts (Chandra, 2004; Iyer and Mani, 2012). The impact that I am looking at is therefore indirect: local leaders, because of their connections, play the role of intermediaries between their caste members and the bureaucrats.

As the caste group of elected leaders is not random, I exploit the institutional reform of 1993 in India that has created a three-tier government system, with elected councils at the village, block and district levels. In addition of giving more power over expenditures to village councils (also called Gram Panchayats), this reform established that the position of village council president (alternatively called "pradhan") had to be reserved for low castes in a certain number of villages, where the number of villages in each State is determined by the proportion that low castes represent in their State population. I use the fact that reservations for the position of president in village councils determine the caste group of the person in power, to look at the impact of having someone from the same caste group as a pradhan on the probability of applying for reservations in the public sector. To take into account the fact that the attribution of political reservations to a specific caste group in a specific village is partially determined by village level characteristics,[2] I focus on intra-village variation across caste groups in the probability of applying for reservations in public employment.

I first document that sharing the same caste group as the pradhan increases households' application for reservations in the public sector. Secondly, I show that this impact is actually restricted to members from the same subcaste (also called jati). There are competitive explanations for this impact, but evidence suggests that this is due to the pradhan actually helping the application of members from his caste group to be processed. The results provided in this paper do not go in the direction of the other channels such as an improved self-confidence or a better access to information.

[2]Whereas every village get political reservations for each caste group at some point in time, the sequence in which villages get political reservations is not random, therefore leading to systematic differences across villages in their observable characteristics (Dunning and Nilekani, 2013). This point will be further explained in section 4.1.

The fact that elected political leaders favor their own group in the distribution of benefits has been well documented in the literature. Looking at the impact of electoral reservations for scheduled castes (SC) and scheduled tribes (ST) at the State level in India, Pande (2003) finds that they have increased redistribution of resources in favor of these groups. The same pattern has been observed at a more local level. Besley et al. (2004) show that the intra-village allocation of public goods is shifted towards low castes when the seat of the pradhan is reserved for low castes. Similarly, Chattopadhyay and Duflo (2004) find that female pradhans distribute public goods which are more relevant to the needs of women.

There is less empirical evidence in the economics literature on how having someone from their group in power changes the behavior of individuals, because they expect different returns from their actions, or because they feel more confident to do so. Among the exceptions, Beaman et al. (2009) find that female reservations for the seat of pradhan leads to an increase in the proportion of female candidates in successive elections.

This paper provides new empirical evidence that emphasizes the importance of caste networks in getting access to State benefits. Whereas previous literature has mainly focused on the impact of political leaders' identity on the distribution of public goods, this paper studies a different kind of public benefit, namely quotas for low castes for jobs in the public sector. This policy is the biggest affirmative action program in the world, but little is known about the mechanisms underlying people's application. This paper therefore sheds some light into what determines or keeps people from taking advantage from reservations in the public sector.

The rest of the paper is organized as follows: section 2 provides some information on the caste system and caste-based affirmative action programs. Section 3 describes the data used and gives some descriptive statistics. Section 4 explains the empirical strategy before showing the main results and exploring the channels. Finally, section 5 concludes.

2 Contextual background

The affirmative action policy in India is due to, and based on, the caste system. To understand the policy studied in this paper, this section provides basic information on the caste system before giving a quick overview on the evolution of reservations in public employment.

2.1 The caste system and the need for reservations

The Indian society is stratified according to a caste system. The organization of the caste system is hierarchical, and is governed by the concept of purity and impurity. People at the top of the hierarchy are considered to be pure, whereas people at the bottom are very impure and therefore "untouchable". The caste system goes hand in hand with a division of occupations across subcaste groups (or jati), where each subcaste group has a traditional occupation. Low caste groups are traditionally doing menial jobs such as scavenging or clothes cleaning. Membership to a specific subcaste (jati) is hereditary and endogomy is practiced, such that it is very difficult, not to say impossible, to escape ones' caste.

Because the caste system in itself does not allow for social mobility, the caste group at the bottom of the social hierarchy, the untouchables (also called dalits), have always been the most economically disadvantaged group. And because of their impurity, they were and still are suffering from discrimination: they had been obliged for centuries to live in specific areas of villages, they were forbidden from using common goods such as wells or places of worship.

At Independence, under the lead of Dr. Ambedkar himself a dalit, affirmative action was written in the Constitution. The untouchables, named as the "Scheduled Castes" (SC) and the indigenous people of India under the name of "Scheduled Tribes" (ST) were provided with respectively 15% and 7.5% quotas in public employment and in higher education institutions. The other low castes, also economically poor and socially backward but not suffering from the stigma of untouchability, were not benefiting from this policy. But a door was left open for them in the Constitution where they were referred as the "Other Backward Classes" (OBC).[3]

2.2 The evolution of reservations for low castes in public employment

Whereas reservations in public employment for SC and ST were quickly implemented and accepted, reservations in public employment for OBC were subjected to several twists and turns and were extremely diverse across States. There has been reservations in the public sector for OBC in the Southern States prior to Independence. During the post-Independence period, several other States adopted reservations for OBC. However, the turning point was the "Mandal Commission" of 1979 which recommended 27% quotas in the public sector for OBC. This recommendation was implemented by the Central Government in 1993. Most of the States followed the

[3]For clarity purposes, in the rest of the paper, when I refer to "caste" it relates to the broad caste group, namely SC, ST and OBC, and when I refer to "jati" it relates to the the hereditary and endogamous subcaste groups which compose each caste group.

Central Government and implemented reservations for OBC in their own services, the most recent State being West Bengal which only gave reservations in the public sector for muslim OBC in 2012.

Currently, there are employment quotas in the public sector (central and State Government) for SC, ST and OBC in every State.

3 Data and descriptive statistics

The data used to conduct this study are from the 2006 round of the ARIS-REDS database from the National Council of Applied Economic Research (NCAER). Since 1971, the NCAER has been conducting household surveys along with village surveys in 259 villages in the 17 major States of India.

Out of the 259 villages where the survey has been conducted, I focus on a subsample for which the jati of the pradhan is available.[4] This is the case for 37 villages situated in three States: Andhra Pradesh, Karnataka and Maharashtra. In those villages, I focus on households from castes benefiting from reservations in public employment, namely SC, ST and OBC. The final sample is composed of 14,081 households. Robustness checks on specifications which do not require the jati of the pradhan are conducted with a broader sample including all the States which have political reservations for OBC[5] and reservations for OBC in State government jobs. The additional States in this sample are Chhatisgarh, Gujarat, Rajasthan, Tamil Nadu and Uttar Pradesh.

The village survey provides the reservation status of the president council position for three periods: current pradhan, previous pradhan and previous to previous pradhan. Table 4.1 summarizes the information for the reduced sample, and for the broader sample used for robustness checks. Two important points are to be noted: first, OBC benefit the most from pradhan's reservation. Each electoral term, more than 50% of the reserved pradhan seats are reserved for OBC. Second, the number of reserved seats constitute almost 50% of the total number of seats for the reduced sample and up to 70% for the reduced sample.

The information about application to reservations for jobs in the public sector

[4]The data do not directly provide the jati of the pradhan, but indicate the name and several socio-demographic indicators along with their electoral score of all the candidates to the seat of pradhan. We also know for all the village members if they were candidate for the seat of pradhan. The jati of the pradhan can be deduced in two specific institutional contexts: when the pradhan is elected among the Gram Panchayat members as in Maharashtra and Karnataka, because the data indicate the jati of panchayat members, or when the Gram Panchayat is composed of only one village, in which case we can match the pradhan to the listing.

[5]Although reservations for SC and ST for the president council position are imposed by the Constitution, reservations for OBC are allowed but not mandatory. Therefore, not all States have implemented political reservations for OBC.

Table 4.1: Villages reservation status

	Reduced sample						Broad sample					
Pradhan position reserved for	current		previous		previous to previous		current		previous		previous to previous	
	nb	%	nb	%	nb	%	nb	%	nb	%	nb	%
SC	3	8%	6	16%	8	22%	14	10%	17	12%	22	17%
ST	4	11%	1	3%	4	11%	9	6%	7	5%	5	4%
OBC	20	54%	18	49%	13	35%	47	33%	40	28%	25	20%
Total Reserved	27	73%	25	68%	25	68%	70	49%	64	45%	52	41%
Non Reserved	10	27%	12	32%	12	32%	73	51%	79	55%	76	59%
Total	37	100%	37	100%	37	100%	143	100%	143	100%	128	100%

The table indicates the number of villages where the seat of pradhan is reserved for SC, ST or OBC, in the reduced sample and the broad sample, for current and previous Gram Panchayat.

is provided in the household survey. The 2006 round is a very peculiar one, because along with the usual questions asked to a sample of households, a complete census of all the households in every village has been conducted. Though the number of questions asked is much smaller than to the sample of households, several questions on reservations have been asked to the households, along with their demographic characteristics. Households were asked about their applications to reservations for jobs in the public sector in the current year of the survey and 10 years prior.

Table 4.2: Households descriptive statistics

	total	mean	sd	min	max
Application to reservation in jobs					
Job reserv app 2006	14081	0.042	0.201	0	1
Job reserv app 1996	14081	0.046	0.208	0	1
Reservations for current pradhan position					
Elect reserv	14081	0.555	0.497	0	1
Elect reserv * caste	14081	0.371	0.483	0	1
SC reserv * SC	14081	0.019	0.136	0	1
ST reserv * ST	14081	0.017	0.129	0	1
OBC reserv * OBC	14081	0.335	0.472	0	1
Households characteristics					
SC household	14081	0.273	0.446	0	1
ST household	14081	0.062	0.240	0	1
OBC household	14081	0.666	0.472	0	1
Age (Years)	14081	44.617	13.090	19	105
Years of schooling	14081	4.012	4.380	0	19
Land owned (in acres)	14081	1.656	8.208	0	650
Household Size	14081	4.576	2.394	0	63
Hindu	14081	0.834	0.372	0	1
Muslim	14081	0.063	0.243	0	1
Christian	14081	0.061	0.239	0	1

The descriptive statistics are provided for the reduced sample.

Table 4.2 provides descriptive statistics for the most important variables for the reduced sample. The mean application rate to reservations in public employment is

rather low and stable over time: 4.6% of the households declare having applied to reservations in 1996 and 4.2% in 2006.[6] Concerning the pradhan, as expected from table 4.1, the proportion of households with a low caste pradhan is very high: 56% of the households in the sample live in a reserved Gram Panchayat. However, only 37% have a pradhan seat reserved for households from their respective caste group. The OBC are the most favored. Among them, 34% of the households benefit from a OBC-reserved pradhan. On the contrary, only 2% of the SC and 2% of the ST have a seat reserved for their caste group.

The demographic characteristics are consistent with what we find in chapter 2 and 3. Low castes are poor and are not very educated. On average their education level is below primary, and they have small landholdings.

4 Results

This section studies if households apply more for reservations in the public sector when they are from the same jati as their pradhan. I first focus on the impact of being from the same caste in subsection 4.1, before looking specifically at the impact of being from the same jati in subsection 4.2. In subsection 4.3 I study what drives the impact measured in the previous subsections.

4.1 Impact of having a pradhan from the same caste

To estimate the impact of having someone from the same caste group as a pradhan on the probability of applying for reservations in government jobs, I take advantage of the reservation system for the position of pradhan. Instead of directly looking at the impact of the caste of the pradhan, which may lead to endogeneity issues,[7] I use the fact that the electoral reservation system imposes that in some villages, only members from a specific caste group can be candidates for the position of pradhan. In the area I am focusing one, the position of Pradhan can be reserved for SC, ST or OBC. When it is not reserved, anyone from any caste group can be a candidate. In practice, low castes are rarely elected when the seat is not reserved. In my sample, only two low caste pradhans have been elected without reservations.

[6]The data do not specify if households got the job after applying. However, for a subsample of households, I am able to know if they were benefited by affirmative action for the job that they have. In mean, 12% of the households who declared having applied for reservations in the public sector also have a member who benefited from affirmative action for his/her job. The percentage reaches 18% for the OBC.

[7]There can be a problem of omitted variables if caste unobservable characteristics, such as connections outside of the village, drive the probability of being elected and of applying for reservations.

The 73rd amendment of the Constitution, which put this policy into place, specifies that "the number of offices of Chairpersons reserved for the Scheduled Castes and Scheduled Tribes in the Panchayats at each level in any State shall bear, as nearly as may be, the same proportion to the total number of such offices in the Panchayats at each level as the population of the Scheduled Castes in the State or of the Scheduled Tribes in the State bears to the total population of the State". For OBC, States do not have to provide reservations for the seat of pradhan but, for those who choose to, the number of seats to be reserved is also determined by law. In Karnataka and Andhra Pradesh, 1/3 of the seats for Chairpersons are to be reserved for OBC, and in Maharashtra it is 27%. Concretely, in most States, the proportion of pradhan seats to be reserved for each caste group is determined at the block level (i.e subdistrict) and is approximately equal to the share that they represent in the population of the subdistrict.[8]

But what matters for identification, is how the reserved seats are attributed to each Gram Panchayat. The 73rd Amendment specifies that the reserved seats have to be allocated by "rotation" across Gram Panchayats, but the order in which each Gram Panchayat get electoral reservation is not specified in the text. In practice, Gram Panchayats are classified according to a specific criterion, which is most of the time the size of the population considered in each Gram Panchayat.[9] In Karnataka, for example, for the attribution of reserved seats for SC, Gram Panchayat are listed in a descending order by the size of their SC population. The sub-district bureaucrats go down the list to allocate the reserved seats (Dunning and Nilekani, 2013). The strategy is the same for OBC reservations in Bihar, except that Gram Panchayats are classified in the descending order of their total population, given that the OBC population is not known. So even if every Gram Panchayat get reservations at a certain point in time, the fact that they get reservation at a specific electoral term is not random, because the order of allocation is determined by a village-level criterion. In a cross-sectional setting, villages with reservations are therefore not similar to villages without reservations and the impact of having a pradhan from the same jati cannot be identified.

To overcome this issue, I use the same identification strategy as Besley et al. (2004), who add village dummies to their specification. The village dummies capture all the village-level unobservables which may be correlated to the reservation status of the village and to the rate of application for reservations in the public sector of a specific caste group, such as the size of their caste group. The impact of having a pradhan from the same caste group on job reservation applications is therefore only

[8]The exact OBC population is not known, because since 1931 the census does not provide figures per caste.

[9]However, the criteria is not known for all States.

identified from within village variation in application. The empirical specification is as follows:

$$Y_{hcv} = \beta_0 + \beta_1 X_{hcv} + \beta_2 CASTE_c + \beta_3 CASTE_c * R_v + \alpha_v + \epsilon_{hcv} \quad (4.1)$$

where Y_{hcv} is equal to one if a member from household h of caste c (namely SC, ST or OBC) in village v applied to reservations for jobs in the public sector. X_{hcv} are household level variables.[10] $CASTE_c$ indicates the caste identity (SC, ST or OBC) of the household. R_v is a set of dummy variables equal to one if the position of pradhan in the village is reserved for SC, ST or OBC. The interaction term between $CASTE_c$ and R_v is therefore a dummy variable equal to one if the pradhan's position is reserved for the caste group of the household. The village dummies α_v control for the unobservables at the village level which may be correlated to the reservation status and ϵ_{hcv} is an error term. The standard errors are clustered at the level of the interaction term, that is to say at the caste in the village level.

The probability of getting reservations is correlated to the size of the SC, ST and OBC population in the block.[11] Some other outcomes at the caste level influencing reservation application may also be correlated to the size of these groups. To take that into account, in some specifications, the regression is also estimated with an interaction term between the caste group and district dummies. Once controlling for this, households from each caste group is similar in reserved and unreserved Gram Panchayat (see table A4.1 in the appendix).

In an alternative specification, I use the panel dimension of the data to look at the impact of having someone from the same jati as a pradhan on application to reservations for jobs. In the survey, households were asked about their application to reservations at the time of the survey (2005-2006) and ten years prior (therefore 1995-1996). For most villages, the data also provide the information on reservations for the seat of pradhan 10 years ago. I am therefore able to study the impact in a panel setting with households fixed effects, where the identification comes from villages where there was a change in the reservation status of the pradhan seat. The econometric specification is as follows:

$$Y_{htcv} = \gamma_0 + \gamma_1 CASTE_c * R_{tv} + T_t + \theta_{tv} + \beta_{hcv} + \xi_{htcv} \quad (4.2)$$

where R_{tv} indicates, as before, if the pradhan's position is reserved for a specific caste at time t in village v, T_t is a time dummy, β_{hcv} is a household fixed effects,

[10] Age and education level of the household head, land owned, household size and religion.

[11] As explained previously, the fact that a village gets reservation for a certain group at a certain point in time depends on its position on the list, which itself depends on the distribution of each caste group in the other villages of the block. The probability of getting reservation consequently depends on the SC, ST and OBC population in the other villages.

which controls for time-invariant unobservables and θ_{tv} controls for time-variant village-level unobservables. I do not control for the caste of the household, because it does not vary over time, and it is absorbed by the household fixed effects. Because the data do not have a real panel structure, I am also not able to control for households' characteristics which vary over time. However, it should not really bias the estimation because the variable of interest is at the caste level. Standard errors are corrected for two-way clustering at the household level and at the level of the interaction term, that is to say at the caste in the village-year level.

4.1.1 Cross-sectionnal evidence

I first begin by looking at cross-sectional evidence of the impact of having someone from the same caste as a pradhan.

Table 4.3 shows the results of the estimation of equation 1. Column 1 looks at the impact of reservations for SC, ST and OBC, and shows that having a pradhan from the same caste group has a positive impact on the probability of members from this caste group to apply for job reservations. Columns 2, 3 and 4 differentiate between the different caste groups. Column 2 and 3 show that having a pradhan who is a SC (column 2) or a ST (column 3) does not increase the probability of applying for reservations of SC or ST.[12] However, having an OBC pradhan for OBC households has a strong and positive impact on their probability of applying for reservations (column 4). The positive impact for OBC is robust to the addition of caste level controls (column 5) and district dummies interacted with caste (column 6), so the impact does not seem to be driven by caste unobservables. As this specification does not use the jati of the pradhan, I can test the impact on the bigger sample. Column 7 shows that the result is robust to a sample change. Column 8 confirms the previous results by using a placebo test. If the impact is driven by caste unobservables at the village level which are correlated with having reservations for the seat of pradhan and the probability of applying for reservations, we would expect that the interaction term between reservations for OBC and the OBC dummy is also positively correlated to the probability of applying for reservations 10 years ago. However this is not the case, the coefficient interaction term is close to 0 and is not significant.

One explanation to the absence of relation between having a SC/ST pradhan and applying to reservations in jobs for SC/ST relates to the power that the pradhan has ones elected. SC/ST pradhans who are elected when the position is reserved for SC/ST are often said to be "puppet candidates" for high castes. Because high castes cannot officially compete in the elections, they help a SC/ST to win such that they

[12]The number of reserved villages for SC or ST in this sample is very small (cf table 4.1), but the same regressions on the broad sample (not shown here) confirm these results.

Table 4.3: OLS estimation of application to reservation for jobs

Dependant Variable:	Application to reservations for jobs in 2005-2006							10 years ago
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age (Years)	0.000779***	0.000782***	0.000783***	0.000781***	0.000789***	0.000812***	0.000997***	0.00107***
	(0.000207)	(0.000206)	(0.000206)	(0.000207)	(0.000206)	(0.000221)	(0.000296)	(0.000354)
Years of schooling	0.00343***	0.00344***	0.00344***	0.00343***	0.00350***	0.00333***	0.0104***	0.00498***
	(0.00118)	(0.00118)	(0.00118)	(0.00118)	(0.00119)	(0.00125)	(0.000895)	(0.00145)
Land owned in acres (in log)	0.0000385	0.000321	0.000318	-0.0000926	0.00138	0.00162	-0.000423	-0.000622
	(0.00257)	(0.00256)	(0.00257)	(0.00258)	(0.00259)	(0.00286)	(0.00339)	(0.00378)
Household Size	0.000580	0.000552	0.000549	0.000583	0.000619	0.000456	0.0199***	0.00184*
	(0.00116)	(0.00116)	(0.00116)	(0.00116)	(0.00116)	(0.00123)	(0.00309)	(0.00109)
ST household	-0.0197**	-0.0179*	-0.0166	-0.0195**	-0.0206**	-0.00914	0.0102	0.0268
	(0.00935)	(0.00989)	(0.0106)	(0.00888)	(0.00829)	(0.0102)	(0.00948)	(0.0173)
OBC household	-0.0310***	-0.0250***	-0.0250***	-0.0365***	-0.0293***	-0.0364***	-0.0463***	0.0157
	(0.00572)	(0.00655)	(0.00638)	(0.00471)	(0.00529)	(0.0127)	(0.0160)	(0.0207)
Elect reserv * caste	0.0144**							
	(0.00684)							
SC reserv * SC		0.000714						
		(0.0155)						
ST reserv * ST			-0.00983					
			(0.0104)					
OBC reserv * OBC				0.0248***	0.0304***	0.0394***	0.0179**	0.00980
				(0.00788)	(0.00763)	(0.00996)	(0.00805)	(0.0143)
Religion dummies	yes	yes	yes	yes	yes	yes	yes	yes
Village dummies	yes	yes	yes	yes	yes	yes	yes	yes
Caste level controls	no	no	no	no	yes	yes	yes	yes
Caste*District dummies	no	no	no	no	no	yes	yes	yes
N	14081	14081	14081	14081	14081	13086	48951	13086
r2	0.0617	0.0614	0.0614	0.0620	0.0631	0.0636	0.158	0.148

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are corrected for clustering at the caste in the village level. The sample in columns 5 and 7 exclude districts where there is only one village for which data are available, to keep some variation when the interaction term between caste and district dummies is added. Column 6 is using the extended sample. The caste level controls are the household level controls aggregated at the caste level, i.e age, education, land owned and household size.

keep control over the council president position. Therefore, the SC/ST candidate elected with the support of high castes does not truly have power over the decisions. OBC on the contrary have more social power than SC/ST and are able to have a candidate not instrumented by high castes. OBC are also often quite numerous at the district or the State level, and OBC Pradhans can rely on their caste networks outside of the village.

Worth noting also is that on average the SC and ST apply more for reservations for administrative jobs. This may be explained by the fact they they got quotas in the public sector much earlier than the OBC.

4.1.2 Panel evidence

Table 4.4 provides further evidence of the robustness of the impact for OBC by using the panel specification. Equation 4.2 is estimated using a fixed-effects procedure, where the unit is the household. In this specification, households who have moved into the village after 1993 are excluded.

Table 4.4: Fixed-effects estimation of application to reservation for jobs

Dependant Variable:	Application to reservations for jobs			
	(1)	(2)	(3)	(4)
2006 year	-0.0160**	-0.0193*	0.0137	0.0118
	(0.00736)	(0.0107)	(0.00884)	(0.00939)
OBC reserv * OBC	0.0259***	0.0373***	0.0131**	0.0151*
	(0.00743)	(0.0130)	(0.00635)	(0.00787)
OBC reserv * OBC * 2006		-0.0211		-0.00400
		(0.0216)		(0.0117)
OBC * 2006		0.0150		0.00339
		(0.0119)		(0.00502)
Village * time dummies	yes	yes	yes	yes
Households FE	yes	yes	yes	yes
N	13553	13553	43594	43594
r2	0.00158	0.00173	0.000140	0.000151

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are corrected for two-way clustering at the household level and caste in the village-year level.

The results are very similar to the ones in table 4.3. Columns 1 and 3 show that having an OBC pradhan has a positive and significant impact on the probability of OBC to apply for reservations, both in the main (column 1) and broader (column 3) samples. Moreover, the estimated magnitude of the impact is very similar to the cross-sectional specifications. Columns 2 and 4 look at if the impact of having

a pradhan from the same caste has changed over time. This is not the case, the interaction term between "having an OBC pradhan for OBC" and the time dummy is not significantly different from 0.

Both specifications (cross-section and panel) show that there is a significant and positive impact of having an OBC pradhan on the probability of OBC households of applying for reservations in jobs. The following section considers a more disaggregated level, the jati level.

4.2 Impact of having a pradhan from the same jati

Section 4.1 underlines that OBC are more likely to apply for reservations in the public sector when the pradhan seat is reserved, and therefore occupied, by OBC. But who is concerned among OBC households? Does the positive impact of having an OBC pradhan concerns any OBC household or is it limited to the close social group of the pradhan, that is to say to people from his jati? Whatever the channel at stake here, the answer is not straightforward. While most of the literature underlines that the jati is still the real group of reference in contemporary India,[13] Dunning (2010) shows that electoral quotas favor intra-caste solidarity. To get some insight into this question, I look at the impact of having a pradhan from the same jati on reservation application. I begin with a simple OLS specification:

$$Y_{hjcv} = \alpha_0 + \alpha_1 X_{hjcv} + \alpha_2 CASTE_c + \alpha_3 E_{jcv} + \delta_v + \omega_{hjcv} \tag{4.3}$$

where E_{jcv} is a dummy variable equal to one if the pradhan is from the same jati as the household and zero otherwise. The equation includes village dummies δ_v such that α_3 is only estimated from intra-village variation. Standard errors are clustered at the level of E_{jcv}, that is to say at the level of the jati in the village.

The OLS results are shown in table 4.5. As expected, the relationship between application to job reservations and having a pradhan of the same jati is positive and significant (column 1). But this correlation may be driven by unobservables, so columns 2 to 5 test the robustness of the relation by controlling for additional factors. Column 2 controls for jati in the village level variables.[14] Interestingly, controlling for jati in the village variables increases the coefficient of the variable which indicates if the pradhan is from the same jati or not, and the coefficient becomes significant at

[13] For recent works on the importance of jatis on diverse outcomes, see for example Damodaran (2008); Munshi (2011) or Munshi and Rosenzweig (2009).

[14] The control variables at the jati in the village level are the mean age of households head in the jati, their mean education level, the mean household size, the mean land owned by the jati in the village and the total number of households in the jati in the village.

Table 4.5: OLS estimation of the impact of having a pradhan from the same jati

Dependant Variable:	Application to reservations for jobs						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Same jati Pradhan	0.0317*	0.0384**	0.0555**	0.0496**	0.0609***	0.0217	
	(0.0175)	(0.0184)	(0.0235)	(0.0248)	(0.0208)	(0.0150)	
Same jati Pradhan * OBC						0.0746*	
						(0.0415)	
Same caste diff jati Pradhan							-0.0192
							(0.0153)
HH controls	yes	yes	yes	yes	yes	yes	yes
Village dummies	yes	yes	yes	yes	yes	yes	yes
Jati in the vil. controls	no	yes	yes	yes	yes	yes	yes
Caste*District dum	no	no	yes	no	no	no	no
Caste*Village dum	no	no	no	yes	no	no	no
Jati dummies	no	no	no	no	yes	yes	yes
N	14081	14081	13086	14081	12867	12867	12867
r2	0.0632	0.0655	0.0681	0.0721	0.129	0.130	0.127

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are corrected for clustering at the jati in the village level. The sample in columns 5 to 8 exclude districts where there is only one village for which data are available, to keep some variation when the interaction term between caste and district dummies is added. .

a 5% level. This result underlines that the probability of having a pradhan from the same jati is correlated to jatis' characteristics, but in a way that biases downward the coefficient on the variable of interest. This is consistent with the fact that households from jatis with a high status (positively correlated to the land owned by the jati) have a lower probability of applying for reservations, because of the stigma attached to reservations, but a higher probability of being elected (Gille, 2013). Column 3 controls for caste at the district level unobservables, by adding an interaction term between caste and district. As there is variation among households from the same caste in the variable of interest, it is also possible to control for caste in the village unobservables (column 4), in which case the impact is solely estimated from within caste in the village variation. Finally column 5 controls for jati unobservables. In all the specifications, having a pradhan from the same jati is positively correlated to applications for reservations in jobs. And once adding jati in the village controls, the coefficient on the jati of the pradhan is quite stable, between 0.04 and 0.06.

The results in section 4.1 also underline that having a pradhan from the same caste has a positive impact only for OBC. To check if the results at the jati level are consistent with these previous results, I look at the heterogeneity of the impact across caste groups in column 6. The result is in line with what was obtained earlier. The fact that having a pradhan from the same jati enhances reservation application is only true for OBC. SC and ST do not seem to be affected by having a pradhan

from their own group. Finally, if the impact of the pradhan only goes through jati networks, we also expect that having a pradhan from the same caste group but a different jati will have no impact. Column 7 considers this question. The coefficient is not significantly different from zero, showing that the increase in OBC reservations application when the pradhan is also an OBC is fully driven by an increase among households from the same jati.

The relationship between having a pradhan from the same jati and application to reservations for jobs is very robust to the addition of controls, and the level of the impact once controlling for some unobservables is much higher than when the impact is estimated at the caste level. It seems to confirm that the impact measured in tables 4.3 and 4.4 is an intra-jati effect. However, having a pradhan from the same jati is not random and may be correlated to characteristics of the jati in the village that cannot be observed and controlled for. Whereas controlling for village dummies in equation 4.1 was enough to control for the potential endogeneity of the *caste* of the pradhan, village dummies do not take into account the selection of someone from a specific *jati* within a given caste. So the OLS regressions in table 4.5 may not accurately estimate the coefficient. Therefore, the impact is now estimated using instrumental variables. The two stages are as follows.

First stage

$$E_{jcv} = \pi_0 + \pi_1 X_{hjcv} + \pi_2 PR_{cv} + \pi_3 (PR_{cv} * PROP_{jcv}) + \pi_4 PROP_{jcv} + \phi_j + \rho_v + \psi_{hjcv} \tag{4.4}$$

Second stage

$$Y_{hjcv} = \lambda_0 + \lambda_1 X_{hjcv} + \lambda_2 \hat{E}_{jcv} + \lambda_3 PROP_{jcv} + \Phi_j + w_v + v_{hjcv} \tag{4.5}$$

The variable of interest, E_{jcv} is instrumented with two instruments. The first instrument is a dummy variable (PR_{cv}) which indicates if there is reservation for the caste group from which the household is from. Namely, the variable is equal to 1 if the household is OBC and the seat of pradhan is reserved for OBC and similarly for SC and ST. As seen in section 4.1, the fact that there is reservation for a specific caste group in the electoral term that is considered here may not be independent of *caste level characteristics* in the block (subdistrict). However, reservations are not correlated with *jatis characteristics in the village*.[15] It is therefore a valid instrument for E_{jcv} and is expected to have a positive impact on E_{jcv}. The second instrument is the

[15]The order in which electoral reservations for OBC are allocated to villages is related to the total size of the OBC population in each village. Jatis' characteristics such as average education or economic strength are not related to the aggregate OBC population.

interaction of the reservation status of the caste group with the proportion in terms of number of households that the jati represents in the village ($PR_{cv} * PROP_{jcv}$).[16] The underlining idea is that the impact of electoral reservations increases with the strength in terms of numbers that the jati has in the village. It is therefore also expected to have a positive impact on the probability of having a pradhan from the same jati. Because in itself the jati proportion may also impact the probability of applying for reservations, I control for $PROP_{jcv}$ in both stages.[17] The second stage is similar to the OLS specification, but the impact of having a pradhan from the same jati is estimated with its predicted value $\hat{E}_{jcv}$. Standard errors are clustered at the jati in the village level.

Table 4.6: IV estimation of the impact of having a pradhan from the same jati

Dependant Variable:	Application to reservations for jobs					
	First stage (1)	Second stage (2)	First stage (3)	Second stage (4)	First stage (5)	Second stage (6)
Same jati pradhan		0.0731***		0.0646***		0.0621**
		(0.0201)		(0.0200)		(0.0257)
Same caste Reserv	0.188***		0.185***		0.271***	
	(0.0665)		(0.0653)		(0.0806)	
Prop Jati * same caste Reserv	0.718***		0.689***		0.461**	
	(0.258)		(0.257)		(0.222)	
Prop Jati	0.0765	-0.0695**	0.211	-0.0721**	0.653***	-0.0318
	(0.205)	(0.0276)	(0.238)	(0.0347)	(0.214)	(0.0394)
HH controls	yes	yes	yes	yes	yes	yes
Village dummies	yes	yes	yes	yes	yes	yes
Jati in the village controls	no	no	yes	yes	yes	yes
Caste*District dummies	no	no	no	no	yes	yes
Jati dummies	yes	yes	yes	yes	yes	yes
N	12867	12867	12867	12867	11992	11992
r2	0.312	0.0127	0.327	0.0131	0.538	0.0171
Instruments F-test		20.93		20.62		17.05
Hansen J-test (p-value)		0.229		0.172		0.223

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are corrected for clustering at the caste in the village level. The sample is only using households from jatis present in at least two villages. In columns 5 and 6, I control for so the sample is further restricted to districts where there are at least two villages to have variation.

Results are shown in table 4.6. The table shows three different specifications, where I allow the controls to vary. In column 1 and 2 the only controls are household level controls, jati dummies and village dummies. Column 3 and 4 additionally

[16] $PROP_{jcv}$ is the number of households from the same jati in the village divided by the total number of households in the village.

[17] The interaction term is not correlated to jatis characteristics, once controlling for the independent effect of $PROP_{jcv}$.

control for jati in the village variables and in column 5 and 6 an interaction term between caste and district is added. In the three first stages, the two instruments have a strong and positive impact on the probability that the household has a pradhan from the same jati. The p-value of the Hansen J-statistics also shows that the exogeneity of the instruments cannot be rejected. Looking at the second stages, having a pradhan from the same jati has a significantly positive impact on the probability of applying for reservations in the three specifications. The impact is quite large, it increases the probability of applying for reservations by 6 to 7 percentage points. It is also similar to what was estimated in the OLS regressions. Interestingly, the jati population share in the village is negatively correlated to reservation application. This is consistent with the results of table 4.3 which showed that the omission of jati level characteristics biases downward the impact of having a pradhan from the same jati.

There is no instrument to estimate the impact of having someone from the same caste group but a different jati. But the fact that the coefficient on the jati of the pradhan is much bigger than when the impact is estimated at the aggregated level indicates that much of the increase in reservation application, if not all, comes from households who are from the same jati as the pradhan.

4.3 What drives the impact?

The fact that having a pradhan from the same jati enhances reservation application can be due to different factors. What inspired that paper is the idea that having a pradhan from the same jati makes it easier to access reserved jobs because the pradhan has political connections that he uses to help households from his jati. But other channels might explain the impact. The pradhan may increase the self-confidence of his jati fellows by showing that a low caste individual is able to hold a prestigious position. Or it can be simply related to information sharing. Having someone from the jati in a position with easier access to information may concretely help to apply for reservations. In this section, I discuss the three channels, and provide evidence that the help to access State institutions is the channel that seems to match best the results.

4.3.1 "Patronage democracy"

The first channel that I am exploring in this section is the one where the pradhan acts as an intermediary to help his jati fellows to get access to reservations. Political science and anthropological literature has emphasized that politicians in India have a great power over the distribution of resources because they control the bureaucracy

(for a more detailed description of the phenomenon, see Chandra, 2004). This "patronage" system makes it complicated for poor people without connections to access State resources. Consequently, in rural India, there is an extensive usage of intermediaries (Manor, 2000; Witsoe, 2012). The pradhan plays an important role in this intermediation. Dunning and Nilekani (2013), in a survey from 512 villages in three Indian States (Karnataka, Rajasthan and Bihar), asked "to whom a hypothetical citizen would most likely turn for help getting access to a government benefit or service". 73% of the respondents answered the council president. They were also asked "who has the most power to provide access to the desired service". To this question 43% identified the president.

However, there may be a discrepancy between what individuals believe that the pradhan can do and what he can do in practice. Indeed, the pradhan can only help his jati fellows to get reservations in the administration if he is himself well connected. While high castes are historically well connected to the State and the political sphere, for low castes accessing power at a very local level it may not be the case. One way pradhans get access to higher levels of bureaucracy is through political parties. Political parties get involved in pradhans' elections by financing their campaigns. In my sample, 2/3 of the pradhans currently in office have been financially supported for their campaigns by a political party. The proportion is the same for OBC pradhans.

I test the hypothesis that the positive impact of having a pradhan from the same jati is due to the pradhan helping households to get access to the State by looking at if the impact differs across pradhans financed by political parties and other pradhans. Namely, I interact the variable of interest with a dummy variable equal to one if the pradhan's campaign was financed by a political party and zero otherwise. If the impact is due to patronage, then we expect the interaction term to be positive and significant. Table 4.7 shows the results for different specifications. Column 1 reproduces the results of the caste-level estimation in OLS; columns 2 and 3 show the OLS estimation of the jati-level estimation, and column 4 shows the IV estimation of the jati-level estimation.[18] In the four specifications, the results are similar: the pradhan only has an impact on reservation applications to jobs when his campaign was financed by a political party. In other words, it is only when the pradhan is connected outside of the village through a political party that OBC tend to apply more for reservations. This result strongly supports the hypothesis of households applying more for reservations when they know that their pradhan can actually help them get the job.

[18]In the IV specification, both the "same jati pradhan" variable and the interaction term are instrumented. The strategy used is the one recommended by Angrist and Pischke (2008). More details are provided in the appendix as well as the two first-stage estimations.

Table 4.7: Party support and job reservation application

Dependant Variable:	Application to reservation for jobs			
	(1)	(2)	(3)	(4)
Estimation Method:	OLS	OLS	OLS	IV
OBC reserv * OBC	-0.0207			
	(0.0159)			
OBC reserv * Party * OBC	0.0665***			
	(0.0170)			
Same jati pradhan		0.00404	0.00529	0.00595
		(0.00697)	(0.0176)	(0.0249)
Same jati pradhan*Party		0.0873**	0.0920***	0.0877***
		(0.0374)	(0.0209)	(0.0236)
HH controls	yes	yes	yes	yes
Village dummies	yes	yes	yes	yes
OBC*Party	yes	no	no	no
Caste in the village controls	yes	no	no	no
Jati in the village controls	no	yes	yes	yes
Jati dummies	no	no	yes	yes
Caste*District dummies	no	no	no	no
N	14081	14081	12867	12867
r2	0.0638	0.0688	0.131	0.0148

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are corrected for clustering at the caste (column 1) or jati (columns 2 to 4) in the village level. The sample in columns 3 and 4 is only using households from jatis present in at least two villages.

But one can also think about other channels playing a role in the impact. In the two following subsections I question the channels of self-confidence and of information sharing.

4.3.2 An increase in self-confidence?

Another reason why the pradhan would enhance applications for reservations of his caste fellows is self-confidence. Psychological literature has underlined that social identity has an impact on performance because people tend to behave according to stereotypes. A reason why they do so is related to self-confidence. In an experiment conducted in India, Hoff and Pandey (2006) confirm that social identity measured by caste has an impact on low caste students' behavior. Students from different caste groups were asked to perform a task. When their caste group was announced, low caste students were doing poorer than when they were anonymous. Discrimination against low castes seems to have been embodied such that low castes behave in the way people expect them to behave. This behavioral impact of discrimination could also lead low castes to negatively self-select themselves in the reservation application

process.

Recent evidence has shown that stereotypes in India are changing and that reservations have a long lasting impact on the way people consider discriminated groups and on the way they consider themselves. Bhavnani (2009) and Beaman et al. (2009) show that quotas for women in Gram Panchayats have changed the probability that they run for office in unreserved panchayat and that they are actually elected. Similarly, reported crimes against women in India have increased with the implementation of reservations for women in Gram Panchayats, because the police is more willing to record those crimes, but also because women feel more confident to report them (Iyer et al., 2012). Chauchard (2010) provides evidence that having experienced a low caste pradhan diminishes discrimination against low castes in the village, and that this is partly due to a change in how their social status is perceived. He further underlines that the way they evaluate their own social status is positively changed. Having a pradhan from the same caste group may therefore change the behavior of low castes towards reservations through a change in the belief of what they can possibly achieve. The pradhan in this context serves as a role model.

It can also be a confidence in institutions mechanism. In a cross-sectional setting, Bros and Borooah (2013) find that low castes groups have a higher confidence in institutions than high castes and attribute it to the policies that low castes can benefit from.

However, several results in this paper do not support the interpretation of the positive impact of having someone from your jati/caste as a pradhan as a "confidence impact". First, as shown in tables 4.3 and 4.5, the impact is restricted to the OBC. Whereas we would expect that the role model channel has principally an impact on castes at the very bottom of the hierarchy, SC/ST pradhans do not have an impact on the application rate to reservations in jobs of their caste fellows. In columns 1, 2 and 3 of table 4.8, I also provide evidence that political reservations did not seem to have changed the way OBC feel discriminated.[19] Finally column 4 in table 4.8 studies the impact of previous reservations in elections for OBC. If the increased probability of applying for reservations in jobs is due to a better self-confidence, then we would expect the effect to last. However, the impact of previous reservations is *negative*. This result underlines that the impact of having an OBC pradhan increases the application rate of OBC during the electoral term, but decreases it afterwards. It shows that the impact is really related to the pradhan currently being in power.

[19]Households were asked if they have been discriminated against because of their caste while seeking job, or if they have been prevented from entering any street or place of worship within the village. The discrimination variable is a dummy variable equal to one if they answered yes to at least one of these questions.

Table 4.8: The self-confidence channel

Dependant Variable:	Discrimination			App job reserv
	(1)	(2)	(3)	(4)
Estimation Method:	OLS	OLS	IV	OLS
OBC reserv * OBC	0.000777			0.0324***
	(0.0181)			(0.00745)
OBC prev Reserv * OBC				-0.0219**
				(0.00943)
Same jati pradhan		-0.0409	-0.00845	
		(0.0275)	(0.0363)	
HH controls	yes	yes	yes	yes
Village dummies	yes	yes	yes	yes
Caste in the village controls	yes	no	no	yes
Jati in the village controls	no	yes	yes	no
Jati dummies	no	yes	yes	no
Caste*District dummies	no	no	no	no
N	14081	12867	12867	14081
r2	0.199	0.221	0.00451	0.0634

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are corrected for clustering at the caste (columns 1 and 4) or jati (columns 2 and 3) in the village level. The sample in columns 3 and 4 is only using households from jatis present in at least two villages. The dependent variable in columns 1 to 3 is a dummy variable equal to one if a member of the household has suffered from discrimination while seeking jobs or has been prevented from entering a street in the village or a place of worship because of her/his caste.

4.3.3 The information channel

Another way the positive impact of having a pradhan from the same caste/jati is the role that the pradhan may play in transferring information to his network. Applying for reservations in jobs is a complicated process. It requires to know how to get a "caste certificate", which proves the caste identity of the applicant, and how to apply for quotas. The pradhan, because he is connected to the outside world, may be a source of information for his caste/jati.

But again, several results do not support the information channel as being the main channel driving the impact. First of all, if the pradhan is a source of information for his network and this is what increases the application to reservations in job, then we expect that the same happens to application to reservations in schools.[20] Applying to reservations in school also requires a caste certificate, as well as a knowledge of the procedure. The only way in which application to reservations in jobs

[20]Low castes also benefit from quotas in higher education institutions. For more information, see Gille (2013).

Table 4.9: The information channel

Dependant Variable:	Application to reserv. in schools			Applications to reserv. for jobs		
	(1)	(2)	(3)	(4)	(5)	(6)
Estimation Method:	OLS	OLS	IV	OLS	OLS	IV
OBC reserv * OBC	0.0103			0.0540***		
	(0.0180)			(0.0141)		
Same jati pradhan		0.0110	-0.0287		0.0715***	0.0750***
		(0.0272)	(0.0346)		(0.0224)	(0.0214)
OBC reserv * OBC * educ				-0.00526**		
				(0.00235)		
OBC * educ				0.00314		
				(0.00211)		
Same jati pradhan * educ					-0.00235	-0.00277
					(0.00150)	(0.00174)
HH controls	yes	yes	yes	yes	yes	yes
Village dummies	yes	yes	yes	yes	yes	yes
Caste in the village controls	yes	no	no	yes	no	no
Jati in the village controls	no	yes	yes	no	yes	yes
Jati dummies	no	yes	yes	no	yes	yes
Caste*District dummies	no	no	no	no	no	no
N	14081	12867	12867	14081	12867	12867
r2	0.291	0.332	0.00751	0.0652	0.129	0.0136

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are corrected for clustering at the caste (columns 1 and 4) or jati (columns 2, 3 and 5, 6) in the village level. The sample in columns 2, 3, 5 and 6 is only using households from jatis present in at least two villages.

and application to reservations in school differ is through the admission process. Whereas access to reservation in schools is given through an anonymous competitive exam, the process to get a reserved job is discretionary. Table 4.9 shows the results of the same regression but the dependent variable is now application to reservations in school. As we can see in columns 1 to 3, there is no impact of having a pradhan from the same jati/caste on application to reservations for schools. One additional argument is that access to information is related to education (see Foster and Rosenzweig, 1995, for India for example). Therefore, if the information channel is driving the impact, we expect that the impact of the pradhan decreases with the education level of the household's head. The rationale behind the second point is that the more the head of the household is educated, the less additional information is required from the pradhan, because the household already has a good access to information. Columns 3 to 6 in table 4.9 show the results of interacting the education level of the household's head with the variable of interest. The impact of having a pradhan from the same caste for OBC seems to diminish with the education level of the household's head (column 4). But this result is not robust when I only consider

households from the same jati as the pradhan (columns 5 and 6). Therefore, the impact of having a pradhan from the same jati on reservations application does not seem to come from a better access to information.

5 Conclusion

The central empirical finding here it that sharing the ethnic group of the village council president increases households' propensity of applying for reservations in the public sector. The use of the electoral quota system, which determines the caste group of the council president, to identify this impact, allows me to rule out that this due to caste specific unobservables. Indeed the interpretation the most consistent with the data is that it is due to a "patronage" system, where the council president helps the members of his jati to access reserved jobs in the public sector. I cannot find empirical support for alternative explanations such as an improved self-confidence or pure information sharing.

These findings have important policy implications. Although India is the country with the largest affirmative action program, many countries implement quotas for discriminated groups. One conclusion from this paper is that the implementation conditions matter a lot for the success of affirmative action. In this specific context, the fact that the hiring process is discretionary and that politicians have important power over bureaucrats does not provide equal opportunities of access to every eligible people. In practice, the reservation policy in the public sector is more beneficial to already connected people (who supposedly need less help) than to the very poor. The way this policy is implemented seems to have consequences conflicting with the original purposes of affirmative action policies.

Conclusion

Although Aristotle was already writing more than 2000 years ago that "Man is by nature a social animal", the fact that human beings develop and behave according to others has been acknowledged in the economic literature only very recently. It is however important to take into account the social character of human beings, notably because it helps understanding behaviors which otherwise would seem irrational. The mechanism of stigma which prevents people from taking advantage of welfare take-up because they care about how others perceive them (and which is explored in chapter 3) illustrates the role that others can play in decision-making.

Taking into account the role of others is even more important in *development economics*, because of the specific institutional context. In particular, developing countries often have very binding social norms which have a strong impact on economic development. Untouchability in India is an example of a social norm which prevents the economic development of a very sizable share of the population. Far from having only a negative role, others can also compensate for problems inherent to developing countries such as market deficiencies. For example, the importance of informal social networks is often underlined in the context of credit markets.

This dissertation aims at providing some evidence of the role of social capital in India, with a focus on its relation with human capital. It explores the relation following two directions. First, in chapter 1 and 2, it looks at the impact of the human capital of others on aggregate and individual productivity. Second, in chapter 3 and 4, it looks at the impact that others have on human capital.

Main findings

The effect of others' human capital is first studied in chapter 1 with a macroeconomic perspective. Using households level data aggregated at the State level in India, I explore how the distribution of education is related to income per capita. I find that a more unequal distribution of education is positively correlated to income per capita. However, the relation varies across States, and depends on the initial level of income. The main result is driven by richer States, whereas a more equal distribution of education seems to be better for poorer States.

The chapter also provides a first insight in the mechanisms that drive the relation. Theoretically the channels are threefold: there may be education spillovers, the returns to education may be non-linear, and/or there may be complementarity between workers education level. The results show that all the three channels seem to play a role in the relation.

Chapter 2 continues the study of the impact of other's human capital on pro-

ductivity, but this time with a microeconomic perspective. More precisely, I study the impact of the education level of members from the same caste in the village on agricultural productivity. I use household level data from the 17 major States of India.

I find that an increase in the average education level of members from the same caste in the village has a positive impact on households' agricultural productivity. The impact solely comes from households in the caste who have agriculture as their main activity. Those who are not farmers have no impact. There is also heterogeneity in the impact depending on the cultivated crop: households cultivating rice benefit less from the education of their caste fellows than households cultivating wheat. This heterogeneity seems to be related to the extent to which learning about a given crop is possible. Furthermore, I am able to show that it is very unlikely that the positive impact of neighbors' education is driven by caste unobservables. These results therefore provide support to the existence of education spillovers in agriculture.

The focus is different in chapters 3 and 4. I am interested in how others impact human capital accumulation, either directly as in chapter 3, or indirectly through others' impact on available opportunities as in chapter 4.

Chapter 3 studies affirmative action programs (also called reservations) for low castes in higher education. The question I am asking in this chapter is how the application to quotas in higher education in India is affected by the fact that affirmative actions programs are highly stigmatized. To study this I use the same database as in chapter 2 and look at the impact of households' social status on application to reservations. I find that application to reservations decreases with social status for the Other Backward Classes, a group of low castes. For the Scheduled Castes and Scheduled Tribes who are at the bottom of the traditional hierarchy, there is no impact of social status. I also find that the negative impact of social status for OBC is stronger for rich households. These results are consistent with a stigma effect in application for reservations in higher education. They are therefore confirming that others matter in education investment, because people seem to care about others' opinion when applying to specific programs.

Chapter 4 is also concerned with affirmative action programs in favor of low castes. But in this chapter the focus is on quotas in public employment. Low castes have reserved jobs in the public sector, but the access is not easy as recruitment is mainly discretionary. I am therefore interested in the role of caste networks in facilitating access to reserved jobs. To do this, I use the same dataset as in chapter

2 and 3, and look at the impact of having someone from the same caste at a local electoral position on the probability of applying for reservations in the public sector.

I find that having someone from the same caste as an elected leader has a positive impact on application for reservations in the public sector. The explanation of this impact which is the most consistent with the data is that local elected leaders help their caste fellows to get access to the jobs thanks to their own connections. The alternative explanations are not consistent with the empirical findings. This chapter therefore shows that peoples' social capital, here the caste group, has an impact on individuals' economic opportunities.

Discussion and ideas for future research

The four chapters of this dissertation are unanimous on the fact that human capital and social capital are strongly interrelated. This thesis therefore provides empirical support to the argument that the study of human capital requires an analysis which incorporates the *others*. In particular, chapters 1 and 2 underline that the role of others should be taken into account when measuring returns to education, because there are complementarities or externalities which can blur the true impact of human capital. But the impact of others on human capital related issues is not always as straightforward as one might think. Chapters 3 and 4 highlight that the impact of social capital on human capital formation can be indirect but nevertheless have important consequences. In the particular context of affirmative action for low castes in India, others matter because people care about what people around them think and because they have an impact on their job opportunities, and therefore on their expected returns to education. More elaborate ways of incorporating the role of others in the economic theory thus need to be invented.

However, incorporating others in the analysis of human capital also brings its share of surprises with results that the idealistic researcher would have preferred not to find. The fact that *equality* of education is associated with lower income per capita as shown in chapter 1 is hard to translate into policy advice. Similarly, finding that education spillovers do not cross caste boundaries in chapter 2 and therefore that low castes who are not well educated have no chance to benefit from more educated people from other castes does not give hope for the improvement of low castes economic status. Even the analysis of affirmative programs directly targeted at improving low castes economic status in chapter 4 shows that State's benefits are captured by castes that are well-connected.

Fortunately, all the findings are not negative. For example chapter 3 shows that one consequence of stigma in affirmative action is to prevent households with a high

status to apply and therefore acts as a safety barrier against the "creamy layers".

Moreover, the exploration of the link between human capital and social capital is far from being complete. This thesis only focuses on the educational component of human capital, but other dimensions of human capital may relate to social capital. For example, it has been shown that stigma is an obstacle to the enrollment in health related programs (Stuber and Kronebusch, 2004). But the relation between health and social capital does not always go in the expected direction. For example, Luke and Munshi (2007) show that low castes invest more in the health of their children than high castes, and that it is related to different returns to human capital across castes. Although their results are related to a specific context in South India, one should not underestimate the fact that social capital may relate differently to different components of human capital.

Similarly, this thesis approximates social capital by the network one is born with, which is the caste in India. However, social capital evolves along the life according to the relations that one engages in. It would be interesting to study how social capital is affected by economic and social mobility. In India for example, sociologists have underlined that scheduled castes having reached high positions thanks to affirmative action programs have cut off all relations with their caste. Bros (2013) shows that although caste is still a strong determinant of individuals' perceived social status, income and education improve how people classify themselves. One can therefore think that approximating the network by the ethnic or caste group is good when groups are homogenous, but may not be such a good proxy when within-group inequalities increase.

Many other things can be done on the relation between human capital and social capital. Among them, further study of affirmative action programs is of great interest. In particular, one interesting question is how the impact of affirmative action programs such as quotas in education depends on preexisting social capital. In India for example, Deshpande and Newman (2007) describe how hard it is for scheduled castes who entered the university thanks to affirmative action programs to find jobs that match their education level and skills. The suggested reasons are that they lack the network and suffer from discrimination. Bertrand et al. (2010) quantify the difference in terms of wages between quotas beneficiaries and the others in India, but to my knowledge no study directly assesses the role played by networks.

It is also essential to understand the general equilibrium effect of affirmative action programs. In India the implementation of quotas in higher education for OBC in 2006 led to an increase in the total number of seats available in universities. If

quotas in education go along with a boom in the supply of educated workers it would inevitably have an impact on the labor market, in terms of wages for example. Although it may be hard estimate to this equilibrium effect empirically, a theoretical model could be a first step towards a better understanding of the global impact of affirmative action programs.

This thesis sheds some light on the relation between human capital and social capital, but many questions remain unexplored and I am merely mentioning few of them. But given the importance that this topic has, notably for the design of educational policies, new research should soon fill this gap.

Appendix

Appendix to Chapter 1

Table A1.1: Value of τ and θ for States in the sample

State	*Year*	θ	τ
A. & Nicober	1983	0.35	0.06
A. & Nicober	1987-1988	0.41	0.08
A. & Nicober	1993-1994	0.48	0.12
A. & Nicober	1999-2000	0.44	0.1
A. & Nicober	2004-2005	0.47	0.17
A. & Nicober	2009-2010	0.82	0.25
Andhra Pradesh	1983	0.39	0.1
Andhra Pradesh	1987-1988	0.45	0.12
Andhra Pradesh	1993-1994	0.57	0.17
Andhra Pradesh	1999-2000	0.75	0.26
Andhra Pradesh	2004-2005	0.74	0.23
Andhra Pradesh	2009-2010	0.96	0.35
Arunachal Pradesh	1987-1988	0.72	0.11
Arunachal Pradesh	1993-1994	0.47	0.1
Arunachal Pradesh	1999-2000	0.63	0.07
Arunachal Pradesh	2004-2005	0.62	0.16
Arunachal Pradesh	2009-2010	0.97	0.21
Assam	1983	0.26	0.04
Assam	1987-1988	0.33	0.08
Assam	1993-1994	0.38	0.07
Assam	1999-2000	0.41	0.09
Assam	2004-2005	0.37	0.08
Assam	2009-2010	0.38	0.08
Bihar	1983	0.37	0.09
Bihar	1987-1988	0.53	0.13
Bihar	1993-1994	0.69	0.2
Bihar	1999-2000	0.73	0.24
Bihar	2004-2005	0.64	0.18
Bihar	2009-2010	0.97	0.2
Chandigarh	1993-1994	1.29	0.76
Chandigarh	1999-2000	1.52	0.83
Chandigarh	2004-2005	2.2	1.41
Chandigarh	2009-2010	2.18	1.85
Delhi	1983	0.95	0.57
Delhi	1987-1988	0.97	0.64
Delhi	1993-1994	1.26	0.97
Delhi	1999-2000	1.29	0.96
Delhi	2004-2005	1.21	0.79
Delhi	2009-2010	1.55	0.85
Goa	1993-1994	1.04	0.21
Goa	1999-2000	1.07	0.29
Goa	2004-2005	1.07	0.31
Goa	2009-2010	1.87	0.31
Gujarat	1983	0.44	0.14
Gujarat	1987-1988	0.44	0.12
Gujarat	1993-1994	0.63	0.18
Gujarat	1999-2000	0.6	0.17
Gujarat	2004-2005	0.63	0.19
Gujarat	2009-2010	0.67	0.23

State	*Year*	θ	τ
Haryana	1983	0.44	0.1
Haryana	1987-1988	0.55	0.14
Haryana	1993-1994	0.74	0.17
Haryana	1999-2000	0.75	0.17
Haryana	2004-2005	0.87	0.25
Haryana	2009-2010	1.12	0.3
Himachal Pradesh	1983	0.39	0.09
Himachal Pradesh	1987-1988	0.41	0.07
Himachal Pradesh	1993-1994	0.6	0.11
Himachal Pradesh	1999-2000	0.78	0.13
Himachal Pradesh	2004-2005	0.85	0.17
Himachal Pradesh	2009-2010	1.13	0.23
J. & Kashmir	1983	0.43	0.13
J. & Kashmir	1987-1988	0.48	0.15
J. & Kashmir	1993-1994	0.62	0.15
J. & Kashmir	1999-2000	0.59	0.14
J. & Kashmir	2004-2005	0.67	0.14
J. & Kashmir	2009-2010	0.95	0.23
Karnataka	1983	0.41	0.09
Karnataka	1987-1988	0.51	0.1
Karnataka	1993-1994	0.67	0.18
Karnataka	1999-2000	0.77	0.21
Karnataka	2004-2005	0.7	0.2
Karnataka	2009-2010	1.03	0.29
Kerala	1983	0.26	0.05
Kerala	1987-1988	0.31	0.07
Kerala	1993-1994	0.42	0.08
Kerala	1999-2000	0.52	0.11
Kerala	2004-2005	0.6	0.13
Kerala	2009-2010	0.72	0.22
Madhya Pradesh	1983	0.28	0.12
Madhya Pradesh	1987-1988	0.32	0.15
Madhya Pradesh	1993-1994	0.58	0.17
Madhya Pradesh	1999-2000	0.47	0.18
Madhya Pradesh	2004-2005	0.46	0.19
Madhya Pradesh	2009-2010	0.68	0.22
Maharastra	1983	0.36	0.1
Maharastra	1987-1988	0.48	0.13
Maharastra	1993-1994	0.55	0.17
Maharastra	1999-2000	0.62	0.19
Maharastra	2004-2005	0.67	0.21
Maharastra	2009-2010	0.95	0.33
Manipur	1983	0.33	0.06
Manipur	1987-1988	0.4	0.13
Manipur	1993-1994	0.74	0.22
Manipur	1999-2000	0.78	0.22
Manipur	2004-2005	0.67	0.21
Manipur	2009-2010	1.14	0.3

State	*Year*	θ	τ
Meghalaya	1983	.22	0.09
Meghalaya	1987-1988	0.29	0.09
Meghalaya	1993-1994	0.24	0.06
Meghalaya	1999-2000	0.28	0.1
Meghalaya	2004-2005	0.26	0.09
Meghalaya	2009-2010	0.34	0.09
Mizoram	1999-2000	0.36	0.06
Mizoram	2004-2005	0.36	0.08
Mizoram	2009-2010	0.32	0.08
Nagaland	1993-1994	0.69	0.13
Nagaland	1999-2000	0.62	0.16
Nagaland	2004-2005	0.74	0.23
Nagaland	2009-2010	0.81	0.21
Orissa	1983	0.27	0.08
Orissa	1987-1988	0.29	0.09
Orissa	1993-1994	0.39	0.12
Orissa	1999-2000	0.51	0.15
Orissa	2004-2005	0.44	0.16
Orissa	2009-2010	0.52	0.17
Pondicherry	1983	0.35	0.06
Pondicherry	1987-1988	0.31	0.09
Pondicherry	1993-1994	0.76	0.21
Pondicherry	1999-2000	0.7	0.2
Pondicherry	2004-2005	0.76	0.23
Pondicherry	2009-2010	1.02	0.53
Punjab	1983	0.47	0.11
Punjab	1987-1988	0.64	0.13
Punjab	1993-1994	0.81	0.17
Punjab	1999-2000	0.97	0.2
Punjab	2004-2005	1.03	0.24
Punjab	2009-2010	1.12	0.27
Rajasthan	1983	0.33	0.11
Rajasthan	1987-1988	0.37	0.16
Rajasthan	1993-1994	0.47	0.19
Rajasthan	1999-2000	0.51	0.19
Rajasthan	2004-2005	0.47	0.18
Rajasthan	2009-2010	0.63	0.25

State	*Year*	θ	τ
Sikkim	1993-1994	0.53	0.08
Sikkim	1999-2000	0.44	0.09
Sikkim	2004-2005	0.41	0.09
Sikkim	2009-2010	0.46	0.17
Tamil Nadu	1983	0.31	0.06
Tamil Nadu	1987-1988	0.37	0.09
Tamil Nadu	1993-1994	0.51	0.13
Tamil Nadu	1999-2000	0.6	0.17
Tamil Nadu	2004-2005	0.62	0.2
Tamil Nadu	2009-2010	0.9	0.26
Tripura	1983	0.38	0.11
Tripura	1987-1988	0.25	0.11
Tripura	1993-1994	0.3	0.1
Tripura	1999-2000	0.3	0.09
Tripura	2004-2005	0.34	0.1
Tripura	2009-2010	0.3	0.1
Uttar Pradesh	1983	0.38	0.14
Uttar Pradesh	1987-1988	0.43	0.16
Uttar Pradesh	1993-1994	0.56	0.2
Uttar Pradesh	1999-2000	0.6	0.22
Uttar Pradesh	2004-2005	0.58	0.21
Uttar Pradesh	2009-2010	0.66	0.25
West Bengal	1983	0.3	0.12
West Bengal	1987-1988	0.28	0.13
West Bengal	1993-1994	0.34	0.16
West Bengal	1999-2000	0.38	0.18
West Bengal	2004-2005	0.38	0.17
West Bengal	2009-2010	0.43	0.17

Table A1.2: Education attainment to years of education conversion table

Educational attainment code	Imputed years of education
Not literate	0
Literate through attending NFEC/AEC, TLC or others	1
Literate, but below primary	3
Primary	5
Middle	8
Secondary	10
Higher secondary	12
Graduate and above	15

Note: NFEC, non-formal education centre; TLC, Total Literacy Campaign; AEC, alternative education centre.

Source: Kingdon and Theopold (2008)

Table A1.3: Inter-State migration in India

	1991		2001					
	Persons	%	Persons	%	Male	%	Female	%
Total population	838,567,936		1,028,610,328		532,156,772		496,453,556	
Population 15+			664,999,516					
By place of birth								
Total migrants	34,241,001	4.1	48,508,633	4.7	22,850,491	4.3	25,658,142	5.2
More than 15+			43,151,813	6.5				
By last place of residence								
Total migrants	32,617,477	3.9	46,321,688	4.5	21,781,996	4.1	24,539,692	4.6
Reasons: Work/Employment			11,317,429	24.4	10,379,479	47.7	937,950	3.8
Business			875,608	1.9	790,719	3.6	84,889	0.3
Education			591,316	1.3	441,264	2.0	150,052	0.6
Marriage			12,869,869	27.8	192,570	0.9	12,677,299	51.7
Moved after birth			1,624,050	3.5	947,755	4.4	676,295	2.8
Moved with HH			11,767,492	25.4	4,801,538	22.0	6,965,954	28.4
Others			7,275,924	15.7	4,228,671	19.4	3,047,253	12.4
Total				100		100		100

Source: Census of India, 1991 and 2001. The percentages represent the percentage of migrants over total population or over the more than 15 years old population, except for the lines indicating the reasons for migration where the percentages represent the percentage of migrants having migrated for a given reason.

APPENDIX

Appendix to Chapter 2

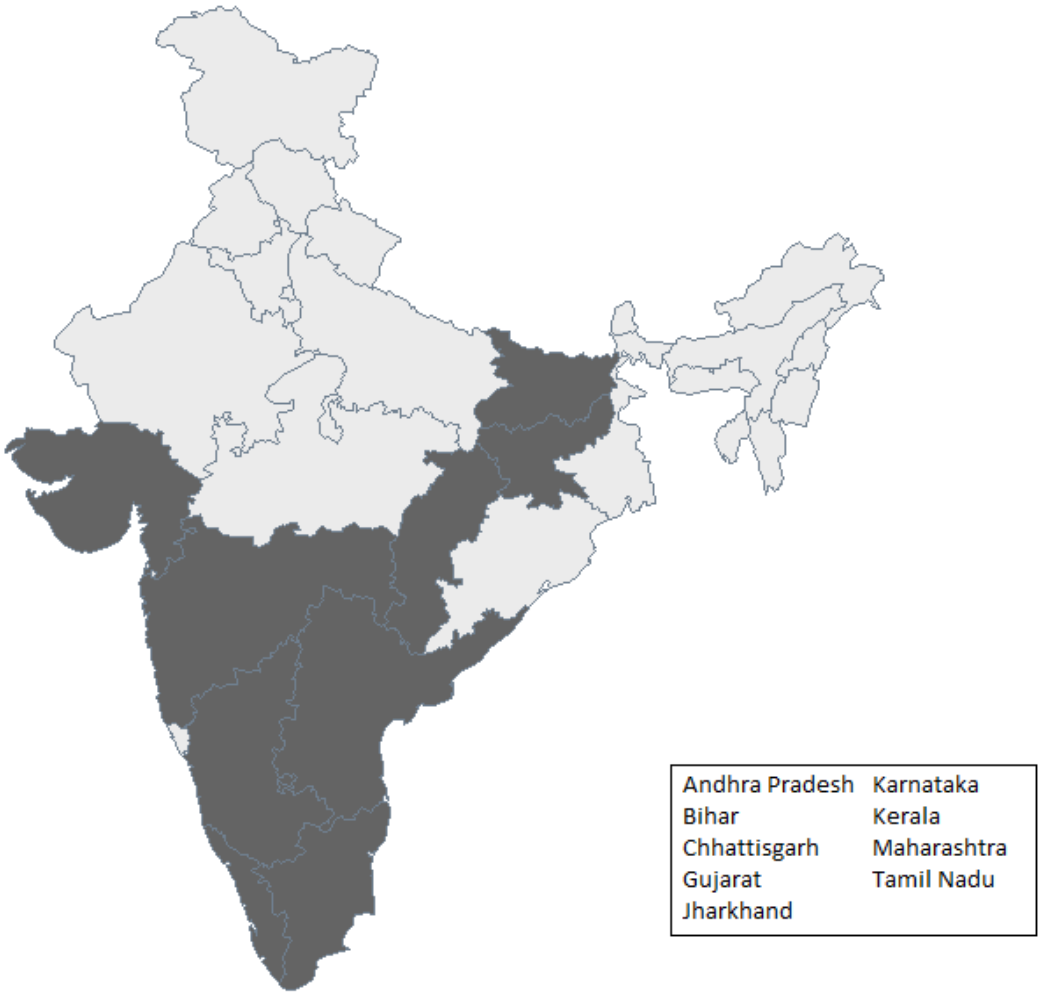
Table A2.1: Low-caste dominated villages

Dependant variable:	Household Productivity			
	(1)	(2)	(3)	(4)
Years of schooling	0.00178	0.00165	0.00147	0.0178***
	(0.00204)	(0.00198)	(0.00198)	(0.00684)
Same occup caste	0.0454***	0.0503***	0.0515***	0.0689***
mean education level	(0.0147)	(0.0151)	(0.0149)	(0.0171)
Head educ * caste				-0.00323**
agr mean education level				(0.00129)
Other occup caste	-0.00117	-0.00118	-0.00161	-0.00124
mean education level	(0.0143)	(0.0153)	(0.0150)	(0.0150)
Low-caste dominated	-0.00905	-0.00327	0.0116	0.0162
village	(0.0521)	(0.0531)	(0.0533)	(0.0528)
HH controls	yes	yes	yes	yes
State dummies	yes	yes	yes	yes
Caste occup controls	yes	yes	yes	yes
Caste oth occup controls	yes	yes	yes	yes
N	4611	4611	4611	4611
r2	0.0342	0.0402	0.0453	0.0495

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, corrected for clustering at the caste in the village level. Caste occup controls and caste ot hoccup controls include the number of people in each group and the mean value for the demographic characteristics (age, proportion of female head, household size).

Appendix to Chapter 3

Figure A3.1: States in the sample



Appendix to Chapter 4

Table A4.1: Differences in means across reservation status

	SC		ST		OBC	
Age	0.180	(0.44)	-0.234	(-0.25)	-0.0576	(-0.22)
Sex	-0.00111	(-0.10)	0.0221	(0.90)	-0.00113	(-0.17)
Years of educ	0.183	(1.34)	-0.236	(-0.84)	-0.0584	(-0.65)
HH size	-0.112	(-1.57)	0.0556	(0.31)	0.0421	(0.85)
Land	0.0114	(0.03)	-0.0593	(-0.31)	-0.000471	(-0.00)
Income	-818.5	(-0.10)	5032.4	(0.74)	-128.9	(-0.02)
Agr activity	0.00180	(0.17)	-0.0352*	(-1.79)	0.00178	(0.25)
Irrig. Land	-0.0875**	(-2.08)	0.229**	(2.01)	0.0196	(0.40)
Observations	3685		820		8945	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table compares the means of various household level variables across villages with different reservation status for the position of pradhan, once controlling for caste*district dummies. For example, column 1 compares variables means in villages where the pradhan position is reserved for SC with variables means in villages where the pradhan is not reserved for SC. Only the differences in means are reported, because once controlling for caste*district dummies, the means in itself are not informative. t statistics are in parentheses.

The strategy to instrument the variable "Having a pradhan of the same jati" and its interaction with the variable indicating if the pradhan's campaign was financed by a political party in section 4.3.1 or with education in section 4.3.3 follows the suggestion in Angrist and Pischke (2008). I provide hereafter the details for the estimation in section 4.3.1, but the strategy is exactly the same in section 4.3.3. I first estimate

$$E_{jcv} = \pi_0 + \pi_1 X_{hjcv} + \pi_2 PR_{cv} + \pi_3 (PR_{cv} * PROP_{jcv}) + \pi_4 PROP_{jcv} + \phi_j + \rho_v + \psi_{hjcv} \tag{A4.1}$$

as in section 4.2. I then use the predicted value, $\hat{E}_{jcv}$ from (A4.1), and its interaction with the dummy indicating if the campaign was financed by a political party $PARTY_v * \hat{E}_{jcv}$ as instruments in the two first-stage estimations of E_{jcv} and $PARTY_v * E_{jcv}$ respectively (A4.2) and (A4.3) below, in a conventional 2SLS procedure.

$$E_{jcv} = \delta_0 + \delta_1 X_{hjcv} + \delta_2 \hat{E}_{jcv} + \delta_3 PARTY_v * \hat{E}_{jcv} + \delta_4 PROP_{jcv} + \gamma_j + \eta_v + \iota_{hjcv} \tag{A4.2}$$

$$PARTY_v * E_{jcv} = \epsilon_0 + \epsilon_1 X_{hjcv} + \epsilon_2 \hat{E}_{jcv} + \epsilon_3 PARTY_v * \hat{E}_{jcv} + \epsilon_4 PROP_{jcv} + \kappa_j + \mu_v + \nu_{hjcv} \tag{A4.3}$$

γ_j and κ_j are jati dummies, η_v and μ_v are village dummies and ι_{hjcv} and ν_{hjcv} are error terms. The estimation results of equations (A4.2) and (A4.3), as well as the first-stage estimations for the IV estimation in section 4.3.3 are reported in table

A4.2. The estimation of (A4.1) is not reported because it is the same as in table 4.6.

Table A4.2: First-stage estimations

	First-stages table 4.3.1		First-stages table 4.3.3	
	(1)	(2)	(3)	(4)
Dependent variable	E_{jcv}	$PARTY_v * E_{jcv}$	E_{jcv}	$educ_{hjcv} * E_{jcv}$
$\hat{E}_{jcv}$	1.115***	0.178	0.983***	-0.256
	(0.175)	(0.114)	(0.154)	(0.693)
$PARTY_v * \hat{E}_{jcv}$	-0.191	0.804***		
	(0.116)	(0.0882)		
$educ_{hjcv} * \hat{E}_{jcv}$			0.00413**	1.021***
			(0.00158)	(0.0265)
HH controls	yes	yes	yes	yes
Village dummies	yes	yes	yes	yes
Jati in the village controls	yes	yes	yes	yes
Jati dummies	yes	yes	yes	yes
N	12867	12867	12867	12867
r2	0.339	0.476	0.329	0.860
Instruments F-test	20.98	49.23	21.67	741.92

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ Robust standard errors are corrected for clustering at the caste (columns 1 and 2) or jati (columns 3 and 4) in the village level. The sample is only using households from jatis present in at least two villages.

References

Acemoglu, Daron and Joshua Angrist, "How large are the social returns to education? Evidence from compulsory schooling laws," Technical Report, National Bureau of Economic Research, Inc December 1999.

Aghion, Philippe, Robin Burgess, Stephen J. Redding, and Fabrizio Zilibotti, "The unequal effects of liberalization: evidence from dismantling the License Raj in India," *American Economic Review*, 2008, *98* (4), 1397–1412.

Aizer, Anna, "Public health insurance, program take-up, and child health," *Review of Economics and Statistics*, 2007, *89* (3), 400–415.

_ **and Janet Currie**, "Networks or neighborhoods? Correlations in the use of publicly-funded maternity care in California," *Journal of Public Economics*, 2004, *88* (12), 2573–2585.

Anderson, Siwan, "Caste as an impediment to trade," *American Economic Journal: Applied Economics*, 2011, *3* (1), 239–263.

_ **, Patrick Francois, and Ashok Kotwal**, "One kind of democracy," Technical Report, University of British Columbia 2011.

Angrist, Joshua D. and Jörn-Steffen Pischke, *Mostly harmless econometrics: an empiricist's companion*, Princeton University Press, 2008.

Appleton, Simon and Arsene Balihuta, "Education and agricultural productivity: evidence from Uganda," *Journal of International Development*, 1996, *8* (3), 415–444.

Arcidiacono, Peter and Sean Nicholson, "Peer effects in medical school," *Journal of Public Economics*, 2005, *89* (2), 327–350.

Asadullah, M. Niaz and S. Rahman, "Farm productivity and efficiency in rural Bangladesh: the role of education revisited," *Applied Economics*, 2009, *41* (1), 17–33.

Baland, Jean-Marie, Catherine Guirkinger, and Charlotte Mali, "Pretending to be poor: borrowing to escape forced solidarity in Cameroon," *Economic Development and Cultural Change*, 2011, *60* (1), 1–16.

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova, "Powerful women: does exposure reduce bias?," *Quarterly Journal of Economics*, 2009, *124* (4), 1497–1540.

Becker, Gary S., *A Treatise on the Family*, Harvard University Press, 1991.

_ , *Human capital a theoretical and empirical analysis, with special reference to education*, 3rd ed., Chicago: The University of Chicago Press, 1993.

Benhabib, Jess and Mark M. Spiegel, "The role of human capital in economic development evidence from aggregate cross-country data," *Journal of Monetary Economics*, October 1994, *34* (2), 143–173.

Berthélemy, Jean-Claude, "To what extent are African education policies propoor?," *Journal of African Economies*, 2006, *15* (3), 434–469.

Bertrand, Marianne, Rema Hanna, and Sendhil Mullainathan, "Affirmative action in education: Evidence from engineering college admissions in India," *Journal of Public Economics*, 2010, *94* (1-2), 16–29.

Besley, Timothy and Robin Burgess, "Land reform, poverty reduction, and growth: evidence from India," *Quarterly Journal of Economics*, 2000, *115* (2), 389–430.

_ **and Stephen Coate**, "Understanding welfare stigma: Taxpayer resentment and statistical discrimination," *Journal of Public Economics*, 1992, *48* (2), 165–183.

_ **, Rohini Pande, Lupin Rahman, and Vijayendra Rao**, "The politics of public good provision: Evidence from Indian local governments," *Journal of the European Economic Association*, 2004, *2* (2-3), 416–426.

Bhavnani, Rikhil R., "Do electoral quotas work after they are withdrawn? Evidence from a natural experiment in India," *American Political Science Review*, 2009, *103* (1), 23–35.

Bitler, Marianne P., Janet Currie, and John Karl Scholz, "WIC eligibility and participation," *Journal of Human Resources*, 2003, *38*, 1139–1179.

Bloch, Francis, Vijayendra Rao, and Sonalde Desai, "Wedding celebrations as conspicuous consumption: signaling social status in rural India," *Journal of Human Resources*, 2004, *39* (3).

Blundell, Richard and Stephen Bond, "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, 1998, *87* (1), 115–143.

Brock, William A. and Steven N. Durlauf, "Interactions-based models," in J.J. Heckman and E.E. Leamer, eds., *Handbook of Econometrics*, Vol. 5 of *Handbook of Econometrics*, Elsevier, 2001, chapter 54, pp. 3297–3380.

Bros, Catherine, "Castes in India: implications of social identity in economics." PhD dissertation, Université Paris 1 2010.

_ , "Social Identity in India: can you earn your way out of caste?," Technical Report 2013.

_ **and Vani K. Borooah**, "Confidence in public bodies, and electoral participation in India," *European Journal of Development Research*, June 2013.

Case, Anne, "Neighborhood influence and technological change," *Regional Science and Urban Economics*, 1992, *22* (3), 491–508.

Cassan, Guilhem, "The impact of positive discrimination in India: Evidence from a natural experiment," 2011. Working paper.

Castelló, Amparo and Rafael Domenech, "Human capital inequality and economic growth: Some new evidence," *The Economic Journal*, 2002, *112* (478), 187–200.

Castelló-Climent, Amparo, "On the distribution of education and democracy," *Journal of Development Economics*, 2008, *87* (2), 179–190.

_ , "Channels through which human capital inequality influences economic growth," *Journal of Human Capital*, 2010, *4* (4), 394–450.

_ , "Inequality and growth in advanced economies: an empirical investigation," *Journal of Economic Inequality*, 2010, *8* (3), 293–321.

Chandra, Kanchan, *Why ethnic parties succeed: Patronage and ethnic head counts in India*, Cambridge University Press, 2004.

Chatelain, Jean-Bernard and Kirsten Ralf, "Spurious regressions and near-multicollinearity, with an application to aid, policies and growth," Documents de travail 12078, Université Paris 1, Centre d'Economie de la Sorbonne November 2012.

Chattopadhyay, Raghabendra and Esther Duflo, "Women as policy makers: Evidence from a randomized policy experiment in India," *Econometrica*, 2004, *72* (5), 1409–1443.

Chauchard, Simon, "Can the experience of political power by a member of a stigmatized group change the nature of day-to-day interpersonal relations? Evidence from rural India," working paper, Darmouth College 2010.

Ciccone, Antonio and Giovanni Peri, "Identifying human-capital externalities: Theory with applications," *Review of Economic Studies*, 2006, *73* (2), 381–412.

Cingano, Federico and Alfonso Rosolia, "People I know: job search and social networks," *Journal of Labor Economics*, 2012, *30* (2), 291–332.

Cohen, Daniel and Marcelo Soto, "Growth and human capital: good data, good results," *Journal of economic growth*, 2007, *12* (1), 51–76.

Colclough, Christopher, Geeta Kingdon, and Harry Patrinos, "The changing pattern of wage returns to education and its implications," *Development Policy Review*, 2010, *28* (6), 733–747.

Coleman, James S., "Social capital in the creation of human capital," *American Journal of Sociology*, 1988, pp. 95–120.

Conley, Timothy G. and Christopher R. Udry, "Learning about a new technology: Pineapple in Ghana," *American Economic Review*, March 2010, *100* (1), 35–69.

Corcoran, Mary et al., "Most workers find jobs through word of mouth.," *Monthly Labor Review*, 1980, *103* (8), 33–35.

Currie, Janet, "The take-up of social benefits," in David Card Alan Auerbach and John Quigley, eds., *Poverty, the Distribution of Income, and Public Policy*, New York: Russell Sage, 2006, pp. 80–148.

Damodaran, Harish, *India's new capitalists: Caste, business, and industry in a modern nation*, Palgrave Macmillan, 2008.

de la Fuente, Angel and Rafael Domènech, "Human capital in growth regressions: how much difference does data quality make?," *Journal of the European Economic Association*, 03 2006, *4* (1), 1–36.

Deliège, Robert, *Les castes en Inde aujourd'hui*, Presses Universitaires de France, June 2004.

Desai, Sonalde and Veena Kulkarni, "Changing educational inequalities in India in the context of affirmative action," *Demography*, 2008, *45* (2), 245–270.

Deshpande, Ashwini and Katherine Newman, "Where the path leads: The role of caste in post-university employment expectations," *Economic and Political Weekly*, 2007, *42* (41), 4133–4140.

Dirks, Nicholas B., *Castes of mind: Colonialism and the making of modern India*, Princeton University Press, 2011.

Dumont, Louis, *Homo hierarchicus; an essay on the caste system.*, University of Chicago Press, 1970.

Dunning, Thad, "Do quotas promote ethnic solidarity? Field and natural experimental evidence from India," working paper, Yale University 2010.

_ **and Janhavi Nilekani**, "Ethnic quotas and political mobilization: Caste, parties, and distribution in Indian village councils," *American Political Science Review*, 2013, pp. 1–22.

Duraisamy, P., "Changes in returns to education in India, 1983-94: by gender, age-cohort and location," *Economics of Education Review*, 2002, *21* (6), 609–622.

Edin, Per-Anders, Peter Fredriksson, and Olof Aslund, "Ethnic enclaves and the economic success of immigrants? Evidence from a natural experiment," *Quarterly Journal of Economics*, 2003, *118* (1), 329–357.

Elbers, Chris and Jan Willem Gunning, "Transitional growth and income inequality: Anything goes," Technical Report, EconWPA 2004.

Eswaran, Mukesh, Bharat Ramaswami, and Wilima Wadhwa, "Status, caste, and the time allocation of women in rural India.," *Economic Development and Cultural Change*, 2013, *61* (2), 311–333.

Falco, Salvatore Di and Erwin Bulte, "A dark side of social capital? Kinship, consumption, and savings," *Journal of Development Studies*, 2011, *47* (8), 1128–1151.

Foster, Andrew D. and Mark R. Rosenzweig, "Learning by doing and learning from others: Human capital and technical change in Agriculture," *Journal of Political Economy*, 1995, *103* (6), 1176–1209.

_ **and** _ , "Technical change and human-capital returns and investments: Evidence from the Green Revolution," *American Economic Review*, 1996, *86* (4), 31–53.

Galanter, Marc, "Who are the Other Backward Classes? An introduction to a constitutional puzzle," *Economic and Political Weekly*, 1978, *13* (43/44), 1812–1828.

Gibbons, Stephen, Olmo Silva, and Felix Weinhardt, "Everybody needs good neighbours? Evidence from students outcomes in England," *The Economic Journal*, 2013.

Gille, Veronique, "Stigma in positive discrimination application: evidence from quotas in education in India," working paper, Universite Paris 1 2013.

Giorgi, Giacomo De, Michele Pellizzari, and Silvia Redaelli, "Be as careful of the company you keep as of the books you read: peer effects in education and on the labor market," Technical Report, National Bureau of Economic Research 2009.

Goux, Dominique and Eric Maurin, "Close neighbours matter: neighbourhood effects on early performance at school," *The Economic Journal*, 2007, *117* (523), 1193–1215.

Gudavarthy, Ajay, "Can we de-stigmatise reservations in India?," *Economic and Political Weekly*, 2012, *47* (52), 55–62.

Hanna, Rema N. and Leigh L. Linden, "Discrimination in grading," *American Economic Journal: Economic Policy*, November 2012, *4* (4), 146–168.

Hanushek, Eric A. and Ludger Woessmann, "The role of cognitive skills in economic development," *Journal of Economic Literature*, 2008, pp. 607–668.

Hayami, Yujiro, "Social capital, human capital and the community mechanism: Toward a conceptual framework for economists," *Journal of Development Studies*, 2009, *45* (1), 96–123.

Heckman, James J. and Jeffrey A. Smith, "The determinants of participation in a social program: Evidence from a prototypical job training program," *Journal of Labor Economics*, 2004, *22* (2), 243–298.

Hnatkovska, Viktoria, Amartya Lahiri, and Sourabh B. Paul, "Breaking the caste barrier: Intergenerational mobility in India," *Journal of Human Resources*, 2013, *48* (2), 435–473.

Hoff, Karla and Priyanka Pandey, "Discrimination, social identity, and durable inequalities," *American Economic Review*, September 2006, *96* (2), 206–211.

Iyer, Lakshmi, Anandi Mani, Prachi Mishra, and Petia Topalova, "The power of political voice: women's political representation and crime in India," *American Economic Journal: Applied Economics*, 2012, *4* (4), 165–193.

_ **and** _ , "Traveling agents: political change and bureaucratic turnover in India," *Review of Economics and Statistics*, 2012, *94* (3), 723–739.

_ **, Tarun Khanna, and Ashutosh Varshney**, "Caste and entrepreneurship in India," *Economic and Political Weekly*, Feb 2013, *48* (06).

Jackson, Matthew O., *Social and economic networks*, Princeton, NJ: Princeton University Press, 2008.

Jaffrelot, Christophe, *Religion, caste, and politics in India*, Columbia University Press, june 2011.

Khamis, Melanie, Nishith Prakash, and Zahra Siddique, "Consumption and social identity: Evidence from India," *Journal of Economic Behavior & Organization*, 2012, *83* (3), 353–371.

Kijima, Yoko, "Caste and tribe inequality: Evidence from India, 1983-1999," *Economic Development and Cultural Change*, January 2006, *54* (2), 369–404.

Kingdon, Geeta, "The quality and efficiency of private and public education: a case-study of urban India," *Oxford Bulletin of Economics and Statistics*, 1996, *58* (1), 57–82.

Kingdon, Geeta Gandhi and Nicolas Theopold, "Do returns to education matter to schooling participation? Evidence from India," *Education Economics*, 2008, *16* (4), 329–350.

Kremer, Michael, "The O-Ring theory of economic development," *Quarterly Journal of Economics*, 1993, *108* (3), 551–575.

Kumbhakar, Subal C. and C. A. Knox Lovell, *Stochastic frontier analysis*, Cambridge University Press, March 2003.

Lanjouw, Peter and Nicholas Stern, *Economic development in Palanpur over five decades*, Oxford University Press, 1998.

Londono, Juan Luis and Nancy Birdsall, "Asset inequality matters: An assessment of the World Bank's approach to poverty reduction," *American Economic Review*, 1997, *87* (2), 32–37.

Lopez, Ramon, Vinod Thomas, and Yan Wang, "Addressing the education puzzle: the distribution of education and economic reform," Technical Report, The World Bank December 1998.

Lucas, Robert Jr., "On the mechanics of economic development," *Journal of Monetary Economics*, 1988, *22* (1), 3–42.

Luke, Nancy and Kaivan Munshi, "Social affiliation and the demand for health services: Caste and child health in South India," *Journal of Development Economics*, 2007, *83* (2), 256–279.

Manor, James, "Small-time political fixers in India's States: "Towel over Armpit"," *Asian Survey*, 2000, *40* (5), 816–835.

Manski, Charles F., "Identification of endogenous social effects: The reflection problem," *Review of Economic Studies*, 1993, *60* (3), 531–542.

Marmaros, David and Bruce Sacerdote, "Peer and social networks in job search," *European Economic Review*, 2002, *46* (4), 870–879.

Marshall, Alfred, *Principles of economics*, Macmillan and co., limited, 1890.

Martins, Pedro S and Jim Y Jin, "Firm-level social returns to education," *Journal of Population Economics*, 2010, *23* (2), 539–558.

Moffitt, Robert, "An economic model of welfare stigma," *American Economic Review*, 1983, *73* (5), 1023–1035.

Moretti, Enrico, "Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data," *Journal of Econometrics*, 2004, *121* (1-2), 175–212.

_ , "Human capital externalities in cities," *Handbook of Regional and Urban Economics*, 2004, *4*, 2243–2291.

_ , "Workers' education, spillovers, and productivity: Evidence from plant-level production functions," *American Economic Review*, 2004, *94* (3), 656–690.

Munshi, Kaivan, "Networks in the modern economy: Mexican migrants in the U. S. labor market," *Quarterly Journal of Economics*, 2003, *118* (2), 549–599.

_ , "Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution," *Journal of Development Economics*, 2004, *73* (1), 185–213.

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

_ **and Jacques Myaux**, "Social norms and the fertility transition," *Journal of Development Economics*, June 2006, *80* (1), 1–38.

_ **and Mark R. Rosenzweig**, "Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy," *American Economic Review*, 2006, *96* (4), 1225–1252.

_ **and** _ , "Why is mobility in India so low? Social insurance, inequality, and growth," NBER Working Papers, National Bureau of Economic Research 2009.

Muralidharan, Karthik and Michael Kremer, "Public and private schools in rural India," Technical Report, Harvard University, Department of Economics, Cambridge, MA 2006.

Pande, Rohini, "Can mandated political representation increase policy influence for disadvantaged minorities? Theory and evidence from India," *American Economic Review*, 2003, *93* (4), 1132–1151.

Park, Jungsoo, "Dispersion of human capital and economic growth," *Journal of Macroeconomics*, 2006, *28* (3), 520–539.

Patacchini, Eleonora and Yves Zenou, "Neighborhood effects and parental involvment in the intergenerational transmition of education," *Journal of Regional Science*, 2011, *51* (5), 987–1013.

_ **and** _ , "Ethnic networks and employment outcomes," *Regional Science and Urban Economics*, 2012, *42* (6), 938–949.

Pritchett, Lant, "Where has all the education gone?," *The World Bank Economic Review*, 2001, *15* (3), 367–391.

Putnam, Robert D, "Tuning in, tuning out: The strange disappearance of social capital in America," *PS-WASHINGTON*, 1995, *28*, 664–664.

Riphahn, Regina T., "Rational poverty or poor rationality? The take-up of social assistance benefits," *Review of Income and Wealth*, 2001, *47* (3), 379–398.

Rodríguez-Pose, Andres and Vassilis Tselios, "Education and income inequality in the regions of the European Union," *Journal of Regional Science*, 2009, *49* (3), 411–437.

Rosenzweig, Mark R. and Andrew D. Foster, "Is there surplus labor in rural India?," Working Papers 991, Economic Growth Center, Yale University October 2010.

Sacerdote, Bruce, "Peer effects in education: How might they work, how big are they and how much do we know thus far?," *Handbook of the Economics of Education*, 2011, *3*, 249–277.

Schipper, Youdi and Johannes G. Hoogeveen, "Which inequality matters? Growth evidence based on small area welfare estimates in Uganda," Technical Report, The World Bank May 2005.

Schultz, Theodore W., "Investment in human capital," *American Economic Review*, 1961, *51* (1), 1–17.

Srinivas, Mysore Narasimhachar, "A Note on Sanskritization and Westernization," *The Far Eastern Quarterly*, 1956, *15* (4), 481–496.

_ , *Caste in modern India and other essays*, Asia Publishing House, 1962.

_ , *Social change in modern India*, New Delhi: Orient Longman, 1987.

Stuber, Jennifer and Karl Kronebusch, "Stigma and other determinants of participation in TANF and Medicaid," *Journal of Policy Analysis and Management*, 2004, *23* (3), 509–530.

Thomas, Vinod, Yan Wang, and Xibo Fan, "Measuring education inequality, Gini coefficients of education," Policy Research Working Paper Series, The World Bank 2001.

Topa, Giorgio, "Social interactions, local spillovers and unemployment," *Review of Economic Studies*, 2001, *68* (2), 261–295.

_ , "Labor markets and referrals," in Alberto Bisin Jess Benhabib and Matthew O. Jackson, eds., *Handbook of Social Economics*, Vol. 1 of *Handbook of Social Economics*, North-Holland, 2011, pp. 1193 – 1221.

Wahba, Jackline and Yves Zenou, "Density, social networks and job search methods: Theory and application to Egypt," *Journal of Development Economics*, 2005, *78* (2), 443–473.

Weir, Sharada and John Knight, "Production externalities of education: Evidence from rural Ethiopia," *Journal of African Economies*, 2007, *16* (1), 134–165.

Witsoe, Jeffrey, "Territorial democracy: Caste, dominance and electoral practice in Postcolonial India," *Political and Legal Anthropology Review*, 2009, *32* (1), 64–83.

_ , "Everyday corruption and the political mediation of the Indian State," *Economic & Political Weekly*, 2012, *47* (6), 47.

Wooldridge, Jeffrey M., *Introductory econometrics: a modern approach*, 2nd ed. ed., Australia: South-Western College Publication, 2003.