

## **Identity Based Policies and Identity Manipulation: Evidence from Colonial Punjab**

Guilhem Cassan

Octobre 2013

**Abstract:** We study identity marker based policies in the context of identity manipulation: we measure the extent of identity manipulation caused by such a policy and describe the mistargeting of public resource this can lead to in the context of colonial Punjab. In particular, we analyze the impact on caste identity manipulation of a caste based legislation, the Punjab Alienation of Land Act (1901). Using original data built from the census of Punjab from 1881 to 1921, we show, using various double and triple differences strategies, that a movement of caste identity manipulation took place in response to the law: ruling out alternative demographic interpretations, we estimate that in 20 years, 7.5% of the population that had an incentive to do so manipulated its caste identity. Using a novel data set on land ownership by caste from 1911 to 1931, we then use an accounting exercise to estimate the amount of mistargeting caused by this manipulation between 3.9% and 8.2% of the total land distributed by the government.

**Keywords:** caste; identity; colonialism; land; public policy.

**JEL Classification:** O12; N45; Q15; H53

**Résumé:** Cette étude porte sur les politiques à destination de groupes d'identité dans un contexte de manipulation d'identité. Nous mesurons l'étendue de la manipulation d'identité causée par de telles politiques et décrivons le mauvais ciblage de ressources publiques en résultant dans le contexte du Punjab colonial. Plus particulièrement, nous analysons l'effet du "Punjab Alienation of Land Act" (1901) sur la manipulation d'identité de caste. A l'aide d'une nouvelle base de données créée à partir des recensements de 1881 à 1921, il est montré qu'en 20 ans, environ 7.5% de la population de la population non bénéficiaire de la loi a manipulé son identité de caste. A l'aide d'une nouvelle base de données sur la propriété terrienne par caste entre 1911 et 1931, il est ensuite montré qu'entre 3.9 et 8.2% des terres distribuées par le gouvernement furent acquises par des personnes ayant manipulé leur identité.

# Identity Based Policies and Identity Manipulation: Evidence from Colonial Punjab.

Guilhem Cassan\*

August 8<sup>th</sup>, 2013

**Abstract:** *We study identity marker based policies in the context of identity manipulation: we measure the extent of identity manipulation caused by such a policy and describe the mistargeting of public resource this can lead to in the context of colonial Punjab. In particular, we analyze the impact on caste identity manipulation of a caste based legislation, the Punjab Alienation of Land Act (1901). Using original data built from the census of Punjab from 1881 to 1921, we show, using various double and triple differences strategies, that a movement of caste identity manipulation took place in response to the law: ruling out alternative demographic interpretations, we estimate that in 20 years, 7.5% of the population that had an incentive to do so manipulated its caste identity. Using a novel data set on land ownership by caste from 1911 to 1931, we then use an accounting exercise to estimate the amount of mistargeting caused by this manipulation between 3.9% and 8.2% of the total land distributed by the government.*

**JEL Classification:** O12; N45; Q15; H53.

**Keywords:** caste; identity; colonialism; land; public policy.

---

\*University of Namur. Email: [guilhem.cassan@fundp.ac.be](mailto:guilhem.cassan@fundp.ac.be). Address: University of Namur, CRED, 8 rue du Rempart de la Vierge, Namur, 5000, Belgium. Tel: +32 (0) 81 72 48 31. Former versions of this paper have circulated under the title “British law and caste identity manipulation: the Punjab Alienation of Land Act”. I am grateful to Véronique Bénéï, Luc Behaghel, Denis Cogneau, Dave Donaldson, Lakshmi Iyer, Eliana La Ferrara, Sylvie Lambert, Ian Talbot, Marie Anne Valfort and Ekaterina Zhuravskaya as well as several seminar and conference participants for helpful comments. I am grateful to the Laboratoire d’Economie Appliquée-INRA and the CEPREMAP for funding the data collection. This paper is produced as part of the project “Actors, Markets, and Institutions in Developing Countries: A micro-empirical approach”, a Marie Curie Initial Training Network (ITN) funded by the European Commission. The maps of this article are made with the Philcarto software: <http://philgeo.club.fr>.

## Introduction

Several countries use identity markers such as skin color or caste as a basis for public policies. Most notably, the United States has an “affirmative action” policy for ethnic minorities, Brazil has quotas in certain universities by skin color, while India has the largest “reservation” program in the world for low castes and tribes (the “Scheduled Castes” and “Scheduled Tribes”, as well as “Other Backward Classes”). Those policies take social identity markers as proxies for economic or social status: since the groups targeted are on average poorer/less educated/discriminated against, having policies based on those markers might be efficient in a context in which obtaining information on, say, income, is costly. However, this type of policy relies on the assumption that those identity markers are easily identifiable and hard to manipulate. This paper questions this assumption, as large mistargeting of policies might be at play if it was to be wrong. In addition, this paper also provides an estimate of the extent of mistargeting linked to identity manipulation. The main contribution of this paper is thus to give the first estimates of such manipulation, convincingly linking a specific caste based policy to caste identity manipulation, as well as describing the extent of mistargeting linked to this manipulation. In order to do so, this paper studies the case of a caste based legislation in colonial Punjab, a major Province of British India. The Indian context provides probably the best case study of such identity based policies, as the Indian identity based policies concern a wide segment of the population, are heavily debated, and are suspected to be diverted (notably) by caste identity manipulation.

A second contribution of this paper is to build a data set allowing to follow castes’ populations for a 40 years period (1881-1921) at the district level. This data enables us to evaluate the causal effect of a caste based policy, the Punjab Alienation of Land Act, on caste identity manipulation in colonial Punjab. By creating an “agricultural castes” category, the membership of which was almost compulsory to acquire land, this law created a very strong incentive to manipulate caste identity in order to claim membership to a caste actually considered as agricultural<sup>1</sup>.

We estimate that 20 years after the law was passed up to 7.5% of the non agricultural caste population had manipulated its caste identity. In addition, an accounting exercise using a novel dataset on land ownership by caste and district from 1911 to 1931, enables us to estimate the extent of mistargeting that this caste identity manipulation led to

---

<sup>1</sup>It is to be noted that similar legislations were also passed after the Independence to protect certain specific groups (for example, Maharashtra’s “Restoration of Lands to Scheduled Tribes Act” of 1974), and which were also suspected to create wide identity manipulation movement (Guha, 2003).

between 3.9 and 8.2% of the land distributed by the Punjabi government during that period.

In order to measure identity manipulation, we will isolate abnormal variations in the evolution of the population of each caste group. The identification strategy used compares the growth rates of the population of the “agricultural” and “non agricultural” castes before and after the law. Thus, it does not rely on a common trend assumption, but on a weaker assumption of stability of the difference in trends. Additional robustness checks implement placebo tests and a triple difference using areas in which the law was not passed (the Princely States of Punjab) as counterfactual. Alternative demographic interpretations of the abnormal evolution of the population of the two caste groups following the Alienation of Land Act are carefully ruled out. In particular, the effects of migration and of demographic transition are demonstrated to be too weak to explain the results. Moreover it is shown that this movement of caste identity manipulation was heterogenous, and was stronger in the districts in which the “agricultural” status also granted additional benefits. A final point of the paper is to describe to what extent the land distributed by the government to “agricultural castes” members was actually obtained by identity manipulators, using an accounting exercise.

A large literature has been focusing on the questions of mistargeting and manipulation in various contexts. [Athreya and Somanathan \(2008\)](#) for example study the case of spatial misallocation of public goods. Mistargeting due to the use of proxy means test has also been recently studied by [Alatas et al. \(Forthcoming\)](#) who show that proxy mean testing can lead to a mistargeting of up to 30% because of the inherent imprecision of proxy means tests. In addition, [Camacho and Conover \(2011\)](#) show that manipulation of the proxies was at play in response to the implementation of a proxy based social program in Colombia. Manipulation of the variables used to identify individuals by policy makers has also been studied in the taxation literature, with the manipulation of taxable income in response to tax rates being one of the standard example ([Saez, 2010](#)). Identity based policies however, are often assumed to be less subject to this type of issues, and the economic literature has so far neglected to deal with this concern for this type of policies. Indeed, identity is often thought as being given at birth, easily observable, and almost not alterable. As such, identity provides a good candidate for the identification of policy beneficiaries, in particular when it is correlated with the economic or social condition, which is often the case, given that many identity groups often suffer specific discriminations.

However, numerous evidence point to identity manipulation in response to identity based policies, both in the Indian context<sup>2</sup> and in other countries<sup>3</sup>. [Francis and Tannuri-Pianto \(2013\)](#) for example study the case of self identification as Black in Brazilian universities following the implementation of quotas for Blacks. However, the extent of identity manipulation, and of the public policy mistargeting that it can lead to has been essentially ignored by the economic literature. This is particularly striking given that several evidence point to the importance of this tendency. Indeed, in India, this movement took such an extent that it was termed as “*demand for disadvantage*” by [Somanathan \(2008\)](#), while [Dudley-Jenkins \(2003\)](#) speaks about “*the proliferation of people claiming to be in one of the various backward categories*”. Hence, there is a strong suspicion that identity manipulation diverts the positive discrimination policies benefits from their beneficiaries on a large scale, but without any measure of how large this manipulation can be and to how much mistargeting it can lead to.

By focusing on the link between public policies and identity manipulation, this paper is linked to the literature on “identity economics” developed by [Akerlof and Kranton \(2000\)](#), who show that the choice of identity can be a crucial economic decision, in a model in which identity yields norms of behavior. It is in particular very close to the theoretical paper of [Caselli and Coleman \(2011\)](#) on ethnic conflict, which notably concludes that “passing” between ethnic groups will take place once an ethnic group dominates the access to resources. Few papers document the formation of ethnic or caste identities, as it requires to follow groups over time in order to ascertain their evolution. For linguistic groups, the study of [Michalopoulos \(2012\)](#) demonstrates the link between ethnic group formation and very long term geographic determinants. [Botticini and Eckstein \(2007\)](#) focus on the role of economic incentives in the conversions from Judaism to Christianity, while [Bodenhorn and Ruebeck \(2003\)](#) document the creation a new racial identities in the southern US states in the *XIX<sup>th</sup>* century. In the Indian context, the only attempt to understand the evolution of the number of caste groups we are aware of is the one by [Ban and Rao \(2007\)](#), which points to a causal impact of the post independence land policy on the number of caste groups, using cross sectional data. While the data used in this paper does not allow an interpretation in terms of caste identity “change” but

---

<sup>2</sup>See for example the recent scandal of fake caste certificates for admission in higher education institutions in Delhi ([The Hindu, 2011](#)), or, for the case of Scheduled Tribes, the work of [Guha \(2003\)](#).

<sup>3</sup>A relatively large sociological literature has documented abnormal variations of ethnic groups in Censuses: see for example ([Lieberson and Waters, 1993](#)) on American Whites or [Nagel \(1995\)](#) on Native American.

only in terms of “manipulation”, it is worth pointing that a wide variety of qualitative work has emphasized the role of British colonial institutions in the construction of caste identity<sup>4</sup>

The first part of the paper presents the law and gives some historical perspective, the second part describes the data, the third part shows the causal impact of the Punjab Alienation of Land Act on caste composition while the fourth part rules out alternative interpretations of the results while the fifth part documents the heterogeneity of caste identity change across districts. The sixth part describes an accounting exercise to estimate the amount of mistargeting that this identity manipulation might have created.

## 1 Historical background

### 1.1 The Punjab Alienation of Land Act

By the end of the 19<sup>th</sup> century, the debt of the landowners had become a concern for the British authorities: *“One of the most significant domestic problem confronting the Indian government [...] was the growing indebtedness of the cultivating classes and a concomitant transfer of landed property [...] to urban moneylenders.”* (Barrier, 1966). This concern was of particular importance in the Province of Punjab, since the Indian army was largely recruiting in the Province (Tai Yong, 2005), and more specifically among the landowning castes<sup>5</sup>. Hence, avoiding rural agitation in that Province was a prime concern and *“...the driving force behind government attempts to find a solution to debt and land transfer was fear for its own position [...]”* (Barrier, 1966), as *“widespread land alienations, many feared, would lead to rural revolt”*(Gilmartin, 1988). The act, which was put in application in June 1901 creates an “agricultural caste” category for which the selling or buying of land was restricted: a member of an agricultural caste could transfer the property of his land (be it by sale or by mortgage) only to an other member of an agricultural caste<sup>6</sup>. Since the members of the agricultural castes were in effect the landowning ones, the members of the non agricultural castes willing to acquire land were almost totally prevented to do so, as only a very small amount of land was

---

<sup>4</sup>See for example Bayly (1999) and Dirks (2001).

<sup>5</sup>Many of which were considered to be “martial races”, see for exemple Vanden Eynder (2011) on the specificities of military recruitment in Punjab.

<sup>6</sup>See Online Appendix 1 for the text of the Act.

available for them to buy<sup>7</sup>

## 1.2 Additional benefits of the “agricultural caste” membership

If the law already provided a substantial protection on the land market, the agricultural caste list was also to be used to identify the beneficiaries of various land related policies. Indeed, the Alienation Act was then reinforced by the Punjab Pre Emption acts of 1905 and 1913 who gave pre-emption rights on land sales to members of agricultural castes. Even more importantly, “ [...] this categorization [...] became the basis for eligibility for land grants in the canal colonies. For land distribution after 1900, the administration did not need to nominate specific groups, but could simply rule that in each selected district the agricultural castes, and those castes alone, were eligible.” (Ali, 1988). Indeed, from the 1880’s on, the colonial administration had dug canals, turning “6 millions acres of desert into one of the richest agricultural regions in Asia” (Talbot, 2007). As the government of Punjab was the owner of most of the land, it was to choose who was to become a “colonist”, and, from 1902 on, allocated the land grants on the basis of the agricultural caste status (Ali, 1988). It can be seen in Figure 1 that the amount of land distributed by the Punjab Government in the canal colonies was massive, and explained by itself the evolution of the land ownership of the agricultural castes in Punjab as a whole.

[Figure 1 about here.]

In a Province in which the vast majority of the population lived in rural areas, being considered as a member of agricultural castes became critical after the enacting of the act, as it became essential to get access to land ownership in the canal colonies, and more generally, to benefit from the protection offered by the status.

## 1.3 Impact on caste identity manipulation

Various reports of the administration mention the different manners in which attempts to avoid the act were made. The first and most obvious one was to lobby the authorities in order to include one’s caste in the list of “agricultural castes”<sup>8</sup>. An other way to evade

---

<sup>7</sup>As underlined by Barrier (1966), the law was successfully enforced: “Sales to non agriculturists ceased after 1901.”. Other references emphasize the impact of the law on the non agricultural castes, such as : “by means of this act moneylenders were practically wiped out of the land market” (Mufakharul Islam, 1995).

<sup>8</sup>Indeed, the number of castes considered as agricultural increased over time. See Online Appendix 4. To rule out the possibility that the increase of agricultural caste is due to increases in the number of



the act was the use of “benami transactions”: using a member of an agricultural caste to buy or mortgage land for a member of a non agricultural caste<sup>9</sup>.

The administrative reports also emphasize caste identity manipulation. Indeed, in the report on the Punjab Alienation of Land Act for the year 1904-1905, it is written: “...members that have acquired money are attempting to get themselves recorded as agricultural tribes with a view to acquiring land...” (Punjab Government, ed, 1906), while for the year 1906-1907, one can see mentions of “...cases of evasive attempt to change tribal designation from a non agricultural to an agricultural tribe in order to defeat the provisions of the Act...” (Punjab Government, ed, 1908). This tendency is reported to be due to individual action, but also sometimes to the mobilization of the caste as a whole: “Frequent cases arise in which application is made by tribes not included in the group notified for the district to have the tribal designation altered to one so included” (Punjab Government, ed, 1909). The movement was so wide that specific recommendations were issued by the administration in order to prevent non “agricultural caste” members to pass as “agricultural”<sup>10</sup>. Those attempts can also be found in various Census reports, that underline a tendency from caste associations to make claims towards the British administration in order to be considered as agricultural. In the Report on the Census of Punjab of 1911 (Kaul, 1912), it can be read : “the introduction of the Punjab Alienation of Land Act [...] has naturally stimulated [...] a tendency to claim an affinity with one or the other of the castes declared by Government as agricultural”<sup>11</sup>.

This takes place in a wider context of caste identity manipulation all across India. Indeed, it has been widely documented (from Ghurye (1932) and Srinivas (1966) to Dirks (2001) and Bayly (1999)) that far from being fixed, the caste system, under the British rule, was evolving under the action of the caste associations (or caste “sabhas”) which were formed in order to “support social advancement” (Assayag, 1995) and to gain access to the economic opportunities created by the British presence<sup>12</sup>.

---

castes considered as such, it is the 1921 list of agricultural castes which is used throughout the paper, any caste being on the agricultural caste list of 1921 being considered as an agricultural caste throughout the period.

<sup>9</sup>For example, in the Report on the Working of the Punjab Alienation of Land Act for 1908 : “What are called benami transactions are reported from most districts. The money lender induces a member of an agricultural tribe [...] to take land on mortgage for the would be borrower” (Punjab Government, ed, 1909).

<sup>10</sup>Those procedures are reproduced in Online Appendix 2.

<sup>11</sup>This claims persisted through time and can also be found in the Report on the Census of Punjab, 1931: “...on the present occasion more than ever before a tendency was noticeable in various localities,[...] to return a higher caste. One of the main reasons was a desire to be included in one of the agricultural tribes [...] to secure exemption from the provisions of the Punjab Alienation of Land Act.” (Khan, 1933)

<sup>12</sup>“the associations began to press for places in the new administrative and educational institutions and for political representation” (Rudolph and Rudolph, 1960)

## 2 Data

### 2.1 Caste Census Data

To estimate the impact of the Punjab Alienation of Land Act on caste identity manipulation, we have collected caste census data from 1881 to 1921. Indeed, every decennial Census collected caste data, which was then tabulated at the district level. It has been widely documented that the Census was part of the mobilization strategies from caste associations, who were very often claiming for new caste names, making the following of each single caste very difficult across time, as both classifications and names might change across time<sup>13</sup>. However, the Punjab Census data is of good quality from 1881 to 1921<sup>14</sup>: using the different Census reports<sup>15</sup> and the Glossary of the Tribes and Castes of the Punjab and North-West Frontier Province (Rose, 1911), we have been able to track the hundreds of changes in classification and names, and merge the newly created caste(s) entries into “caste groups” that are comparable across Censuses<sup>16</sup> and thus building what we believe to be the first dataset following caste groups demography over time at such a disaggregate level<sup>17</sup>. However, the various modifications of district borders and the partition of the North West Frontier Province from Punjab in 1901 as well as the creation of the Delhi Province in 1911 have led us to leave aside some districts while merging some others, in order to assure their comparability over time (see Figure 2).

[Figure 2 about here.]

Overall, we are able to follow 86 caste groups, 24 of which are agricultural<sup>18</sup> in at least one district, and which represent from 97.7% to 99% of the population of the 33

---

<sup>13</sup>See Conlon (1981) on that matter.

<sup>14</sup>The reason why we do not use the 1871 and 1931 Census is that they do not report castes at such a fine level as the other years, thus not allowing us to track all castes for those years.

<sup>15</sup>In particular, the Census report of 1911 contains an “Ethnographic glossary of castes” listing many caste synonyms.

<sup>16</sup>See Online Appendix 3 for the details of this grouping and its justification.

<sup>17</sup>Both geographically fine, at the district level, and fine at the caste level, since we follow caste groups, and not only “scheduled castes” and “scheduled tribes” as is usually the case in most datasets.

<sup>18</sup>More castes and tribes were actually considered as agricultural, but in order to be able to track them over time, we had to merge them either with other agricultural castes, or with non agricultural ones (which bias the results downward). We code as “agricultural” all caste or tribe entered in the “agricultural tribes” list before 1921. The source used for this classification is Lal (1937), see Online Appendix 4 for the list of agricultural castes.

districts and states we are tracking over time, which themselves contain 88% of the population of the Province of Punjab. We have thus built a district level panel of caste composition allowing to study at a very fine geographical level the response of caste groups to the Punjab Alienation of Land Act. It is to be noted that Punjab became part of the British Raj in 1849, so the data used in this paper has been collected more than 30 years after the conquest of the Province, when the British administration had already acquired a good knowledge of the local conditions. Hence, the caste identity manipulation estimated in this paper is a good proxy for the caste identity manipulation faced by the British bureaucracy, and if anything, is a lower bound, as one as no incentive to pass as a member of an “agricultural” caste in front of the Census administration, since there was no link between the Census and the administration in charge of the Alienation of Land Act.

## 2.2 Descriptive Statistics

The whole Province of Punjab had a population of 24.4 million in 1901, for an area of 354,634 square kilometers. It corresponds to the contemporary States of Punjab (Pakistan), Punjab (India), Himachal Pradesh (India) and Haryana (India). As for the rest of India, it was not entirely administered by the British, since some areas, the Princely States, were under the rule of local Princes, and as such, were not subject to British law (see [Iyer \(2010\)](#) for more details, and [Figure 3](#) for their localization), the population of the Princely states was 4.4 millions, thus leaving 19.9 millions under direct British rule. Hence, most of this paper only deals with the districts directly ruled by the British administration, even if the Princely States are also used as a control group in certain specifications.

[Figure 3 about here.]

[Table 1 about here.]

The Province of Punjab was essentially rural, with 89% of the population living in a rural area<sup>19</sup>, hence most of its population was directly concerned by the act, while the urban population was also affected if it wanted to own land.

Within the British districts, the population was roughly cut in half between agricultural castes and non agricultural castes, as can be seen in [Figure 4](#). However, the

---

<sup>19</sup>The Urban population is defined as “(1) Every municipality of whatever size.(2) All civil lines not included within municipal limits.(3) Every cantonment.(4) Every other continuous collection of houses, permanently inhabited by not less than 5.000 persons, which the Provincial Superintendent may decide to treat as a town for census purposes.” (Report on the Census of Punjab, 1901 ([Risley, 1903](#)))

differential evolution of the populations of the two groups is very striking: while the agricultural castes were growing slower before 1901, after the enacting of the law, the growth rate differential with the non agricultural castes is reversed. One can also note that from 1901 on, the overall population did not increase as fast as it did before. This is due to several demographic shocks affecting the Province that will be discussed later on.

[Figure 4 about here.]

Overall, the evolution of the two caste groups' population is coherent with the Alienation Act leading to a movement of caste identity manipulation: the trends in the evolution of the “agricultural” and “non agricultural” castes' population seem to have been inverted after 1901.

### 3 Empirical Approach

#### 3.1 Non agricultural castes as a “control” group

The fact that only certain castes were considered as “agricultural” by the act does not allow to use a simple double difference strategy, as the common trend assumption can not be made here. Indeed, as “agricultural castes” were not randomly selected, and exhibit systematic differences from “non agricultural” ones: they are on average larger (Table 2) and their growth rate before 1901 is on average slower than that of non agricultural castes, as illustrated in Figure 4. Hence, running a difference in difference in levels of population would lead to a underestimation of the amount of identity manipulation.

[Table 2 about here.]

To account for this, we will compare the growth rates of the populations of the two caste groups before and after the law. In this case, the identification relies on the much weaker hypothesis that the differences in the growth rates of the population of agricultural castes versus non agricultural castes before and after 1901 would have remained stable in the absence of the law, and not that their growth rates themselves were similar.

Hence, we will run regressions of the form :

$$\begin{aligned} \ln(pop_{idt}) - \ln(pop_{idt-1}) = & \text{constant} + \beta agr_i + \gamma post1901_t \\ & + \delta agr_i * post1901_t + \eta X_{dt} + \epsilon_{it} \end{aligned} \quad (1)$$

The growth rate (approximated by the difference in log) of the population of caste  $i$  in district  $d$  (if the regression is at the district level) during each of the four decades  $t$  is thus regressed on  $agr_i$  a dummy indicating whether caste  $i$  will be considered as an agricultural caste after 1901,  $post1901_t$  a dummy taking a value of 1 when the decade is in the 1901-1921 interval and 0 in the 1881-1901 interval, and  $X_{dt}$  a set of district dummies, and district dummies interacted with decade dummies, to control for any possible district specific change in trend (if the regression is at the district level). As small castes might tend to have more extreme growth rates, all regressions are weighted by the population of the caste in 1881<sup>20</sup>

We use two main specifications of this regression. In specification 1, we regress the growth rate of the caste population at the British Punjab level, while in specification 2 and 3, we regress it at each British district level, which allows us to control in specification 3 for any district specific change in trend that might have been driving the results (for example, a district with a higher than average share of non agricultural castes that would have been less exposed to some negative demographic shock).

It is to be noted that while this method is formally identical to a double difference on the growth rates, this is not exactly what it is, in the sense that the non agricultural castes are also treated: the growth rate of their population will be affected by the law, since some of their members will pass to the other caste group. The interpretation of the coefficient on  $post1901_t$  is thus different from a classical difference in difference as not only does it present the counterfactual change in growth rate of the “agricultural castes” after 1901 absent the law, but also the decrease in the growth rate of the non agricultural caste population after 1901 in response to the act. However, this does not affect the interpretation of the coefficient of interest,  $post1901*agricultural$ , which is the evolution in the difference of the growth rate of the population of agricultural versus non agricultural castes. This also might raise concerns about the way in which the residuals are correlated. Indeed, as the same units are observed several times before and after treatment, serial correlation is an obvious issue, indicating the need to cluster at the caste level (Duflo and Mullainathan, 2004). However, since there is passing from caste to caste, an across caste correlation of the residuals is not to be excluded, and hence clustering at the district\*decade level would be indicated. In order to take into account those two concerns simultaneously, all the regressions made at the district level are two way clustered at the district\*decade and caste levels (Cameron et al., 2011).

[Table 3 about here.]

---

<sup>20</sup>The results are robust to the choice of an other year.

As can be seen in Table 3 the very precisely estimated coefficient on the interaction between agricultural and post1901 is positive and significant in all specifications. Thus, the average agricultural caste saw its difference in growth rate with the average non agricultural caste increase by around 8 percentage points for every decade after 1901. This points to a very strong effect of the Punjab Alienation of Land Act on caste identity manipulation: while the non agricultural caste were growing slightly faster before 1901, after that date, the growth rate differential was completely changed. As can be seen in specification 3, this effect is not driven by an outlier district, as the results remain robust to including an interaction term of decade dummies with district dummies, controlling for any district specific change in demography. However, the negative coefficient on the post1901 dummy might underline the fact that after 1901, the average caste tended to see its population increase on average less, leading us to suspect the existence of some demographic shocks that would affect Punjab after 1901.

### 3.2 Demographic shocks

Hence, one might argue that the results obtained are not due to caste identity manipulation, but solely that those demographic shocks affected more non agricultural castes than agricultural ones. And as a matter of fact, the 1901-1921 period faced various episodes of epidemic, with plague, malaria and influenza killing millions.

To account for this, we will resort to the neighboring Princely states, which were not subject to the British legislation (see Iyer (2010) for more details). Arguably, the States of Punjab faced the same epidemics as the British districts, due to their close proximity, but were not concerned by the Punjab Alienation of Land Act, thus providing a counterfactual that allows us to control for the demographic shocks of the period. The castes located in the Princely States of Punjab are indeed similar to the castes of the British districts, are subject to the same epidemics, but are not concerned by the agricultural/non agricultural castes categories created by the law. Hence, if the variation in caste groups populations observed in British Punjab were to be attributed to the Alienation of Land Act, we would expect the Princely States caste groups not to exhibit any specific change around 1901 as was the case in British Punjab. Indeed, we can see in Figure 5 that the populations of both agricultural and non agricultural tribes exhibit relatively similar trends throughout the period in the Princely States. It is to be noted that the Princely States can not be considered as perfect counterfactuals: as the work of Iyer (2010) has shown, Princely States and British districts differed in systematic ways. However, as the identification strategy used does not rely on a common trend

assumption, for the Princely States not to be a valid counterfactual, one would need to argue that the difference in the growth's rates of agricultural and non agricultural castes populations would react differently in the British district and in the Princely States when confronted to similar shocks.

[Figure 5 about here.]

We will thus estimate regressions of the form :

$$\begin{aligned} \ln(pop_{idt}) - \ln(pop_{idt-1}) = & constant + \beta agr_i + \gamma post1901_t + \delta agr_i * post1901_t \\ & + \rho british_d * agr_i + \pi agr_i * post1901_t * british_d + \eta X_{dt} + \epsilon_{itd} \end{aligned} \quad (2)$$

With the same notation as in Model 1 and  $british_d$  a dummy indicating that district  $d$  is a British district, and  $X_{dt}$  either the interaction of  $british_d$  and  $post1901$  (and their main effects) or the interaction of district dummies and decade dummies (and main effects), in some district level specifications.

One has to keep in mind that this identification strategy tends to bias the coefficient downwards, as it assumes that the law had no impact in the Princely States, which is far from being obvious: a person living in a Princely State but near a border with a British district would be affected by the law if it were to try to buy some land just on the other side of the border, and would thus face very similar incentives to that faced by a British district inhabitant.

We can see in Table 4 that the coefficient on  $post1901 * british * agricultural$  is significant and positive in all the specifications, and is even larger than in the within British districts specifications. Hence, it appears that the tendency for agricultural castes to grow relatively faster than the non agricultural ones after 1901 than before is specific to British districts, the districts where the law was passed. This confirms the fact that the results obtained are not driven by asymmetric demographic shocks but by the impact of the law itself. Even more so, the estimated impact of the Act with this identification strategy is roughly consistent with the results of the first one, with an implied impact ranging between 10 to 13 percentage points.

[Table 4 about here.]

Appendix A proposes an additional test, allowing the coefficient on  $post1901 * agricultural$  to vary according to the level of exposition to the diseases of the district (with no significant differences across districts), while Appendix B also performs additional placebo tests, showing that the only year during which the growth rates of agricultural and non

agricultural castes diverged was around 1901. Hence, it seems now clear that the law had a causal impact on the caste composition of Punjab. However, as has been underlined earlier, since the identification strategy is not exactly a double difference, the coefficients can not directly be used to infer the number of individuals who manipulated their caste identity in response to the law. The next subsection will deal with this issue.

### 3.3 Evaluating the number of caste identity switchers

Indeed, evaluating the number of caste switchers from the regression tables is not straightforward. As a matter of fact, the growth rate of caste  $i$  during decade  $d$  can be written (absent the error term):

$$g_0\mathbb{1}(post_d) + g_1\mathbb{1}(post_d)\mathbb{1}(agricultural_i) + g_2\mathbb{1}(post_d)\mathbb{1}(nonagricultural_i) + g_3\mathbb{1}(agricultural_i) + g_4\mathbb{1}(nonagricultural_i) \quad (3)$$

With  $g_0$  the change in the growth rate of all castes after 1901,  $g_1$  (resp.  $g_2$ ) the increase (resp. decrease) in the growth rate of agricultural castes (resp. non agricultural) due to the entry (resp. exit) of caste members from non agricultural castes (resp. to agricultural castes),  $g_3$  (resp.  $g_4$ ) the growth rate of agricultural (resp. non agricultural) castes throughout the periods. As  $\mathbb{1}(agricultural_i) = 1 - \mathbb{1}(nonagricultural_i)$ , the former expression simplifies to:

$$(g_0 + g_2)\mathbb{1}(post_d) + (g_1 - g_2)\mathbb{1}(post_d)\mathbb{1}(agricultural_i) + (g_3 - g_4)\mathbb{1}(agricultural_i) + g_4 \quad (4)$$

Hence, both the coefficients on  $post$  and  $post * agricultural$  contain  $g_2$ , the change in the growth rate of the non agricultural caste due to passing to agricultural castes. Thus, the coefficient on  $post * agricultural$  can not be directly used to compute the number of persons passing from non agricultural to agricultural castes. However, we know that  $g_1 * pop\_agricultural_d = -g_2 * pop\_nonagricultural_d$  where  $pop\_agricultural_d$  is the population of the agricultural castes at the beginning of decade  $d$  (and respectively for  $pop\_nonagricultural_d$ ): the number of persons leaving the non agricultural castes must equal the number of persons entering the agricultural castes. Finding the values of  $g_0$ ,  $g_1$  and  $g_2$  is then just a matter of resolving an equation system of three equations with three unknown, resolved in Online Appendix 6. Hence, in 1921, there was a total of 656 177 caste identity switchers, which represents 3.8% of the total British Punjab population or 7.5% of the non agricultural caste population.



## 4 Ruling out alternative interpretations

However, it is unclear how the impact of the law should be interpreted: while the qualitative evidence taken from the Census and administrative reports point to caste identity manipulation, we can not yet rule out other interpretations. In this section, we propose to rule out the two main alternative interpretations of the results: first, that the caste composition of migration might have changed in reaction to the act, and second, that the law created better living conditions for the castes it targeted, hence making those castes enter the demographic transition earlier than the other ones.

### 4.1 Change in the caste composition of migration

Indeed, a very plausible interpretation would be that the results are entirely driven by migration: after the law was passed, members of the castes that would be considered as agricultural in the British districts of Punjab faced an incentive to migrate from their place of origin to a British district of Punjab in order to benefit from the status that the law gave them. The symmetric case is more probable, with members of non agricultural castes leaving British Punjab, to find places in which they are allowed to buy land. To rule out this interpretation, we use the birth place statistics of the Census<sup>21</sup>. At around 5%, the share of the persons not born in a British district residing in such a district is relatively small. However, what can not be seen is whether the caste composition of migration has changed after 1901 towards more arrivals of members of agricultural castes. Moreover, the birthplace data does not allow to know when exactly the migration took place while it is the migration taking place between 1901 and 1921 which is likely to bias the results. In order to compute the migration taking place during this period, we would need to know how many of the persons not born in a British district and residing in such a district in 1901 were still present in 1911 or 1921 (and symmetrically for the emigrants from British districts). The Vital Statistics of India provide yearly district level data on the number of death in Punjab. Thus, we can compute the migration taking place between the Princely States and British districts of Punjab between 1901 and 1921 as:

$$Imm_{jdt} = pop_{jd}^x - pop_{jd}^{x-10} * \prod_{i=x-10}^x survivalrate_{id}$$

With  $Imm_{jdt}$  the number of immigrants coming from district  $j$  to district  $d$  during

---

<sup>21</sup>We thank Dave Donaldson for having given us access to this data.

decade  $t$  (with  $j$  a Princely State or a district located outside of Punjab and  $d$  a British district of Punjab, and  $t$  either 1901-1911 or 1911-1921),  $pop_{jd}^x$  and  $pop_{jd}^{x-10}$  the number of persons born in district  $j$  and enumerated in district  $d$  in year  $x=1911$  or  $1921$ , and  $survivalrate_{id}$  is 1 minus the death rate of district  $d$  in year  $i$ . Symmetrically, the number of emigrants from British districts to non British districts<sup>22</sup> can be calculated.

To check if migration is indeed driving the results, we then recompute the variations of population of each caste group, but this time subtracting the population of immigrants from the population of agricultural castes and by adding the population of emigrants, assuming that the migrants are distributed across the different castes proportionally to their respective sizes<sup>23</sup>. Hence, we make the extreme assumption that after 1901 all immigrants into British Punjab are agricultural castes members while all emigrants leaving British Punjab are non agricultural castes members. Reproducing the first identification strategy (described in Model 1), but this time removing any influence that migration might have had, we can check if the results obtained were or not only driven by migration. Table 5 reports the results and shows that even under the extreme assumption that all post 1901 immigrants to British Punjab are members of agricultural castes (and symmetrically for emigrants), the coefficient on post1901\*agricultural is still positive and significant.

[Table 5 about here.]

## 4.2 Better economic conditions for agricultural castes due to the Act

An other straightforward interpretation could be demographic transition. To assess the validity of this interpretation, we will look at the age composition of each type of caste. If this interpretation was to be true, the structure of the age pyramid would appear to be different for each type of caste. Two scenarios (and/or any combination of the two) can be thought of: one in which the fertility rate of the agricultural castes increases dramatically (or the death rates of the children decreases dramatically), and one in which the death rates of the older population of agricultural castes decreases. The first scenario

<sup>22</sup>As the district of birth of Punjabis enumerated outside of Punjab is not known, the Punjabis emigrants are allocated to each district proportionally to the district's share in the total population. Also, as the data on death rate is not available outside of British Punjab, the death rate of each Punjab's Princely State is assumed to be equal to British Punjab's average death rate, while outside of Punjab, it is assumed that all persons born in Punjab and enumerated after 1901 outside Punjab migrated after 1901.

<sup>23</sup>ie. we subtract  $x\%$  of the population of a district's immigrants from the population of an agricultural caste representing  $x\%$  of the district's agricultural tribes population, the opposite exercise being done for emigrants and non agricultural castes.

would result in the base of the age pyramid being relatively larger for agricultural castes, the second scenario with the top of the pyramid being relatively larger for agricultural castes.

The Census reports give the composition by age of certain castes for the whole Province for the year 1911<sup>24</sup>. The 58 castes for which this information is reported represent 90% of the total population of the Province of Punjab in 1911. One can see in Figures 6 that the age structures of the two caste groups appear to be very similar, pointing to the fact that the law seems not to have had a large impact on the fertility and infant death rates of the agricultural castes. Indeed, the share of children under the age of 11 (born from 1900 to 1911) in the agricultural castes is slightly lower than the share of the same age group in the non agricultural castes, thus invalidating the fertility rate/decrease in child death rate scenario to explain the increase in agricultural castes' share in the population. However, the top of the pyramid is slightly larger for agricultural castes than it is for non agricultural castes: with 22.93% of the agricultural castes population against 22.27% of the non agricultural castes population. A small back of the envelope calculation is sufficient to rule out any major role of this difference in the evolution of the share of the agricultural castes. Indeed, if the share of the persons aged 39 years old or more in the agricultural castes population was to be 22.27%, it means that this age group's population should be 104,000 smaller than it is<sup>25</sup>. Overall, this would mean a decrease of the total agricultural caste population for which the age data is available of 0.85%, or a decrease of the share of the agricultural castes total population in 1911 of 0.21 percentage points. However, the share of the agricultural castes population had increased by 1.75% between 1901 and 1911. The scenario of the Alienation of Land Act reducing the elderly death rate is thus not sufficient to explain the whole evolution of the agricultural castes' population.

[Figure 6 about here.]

It thus appears that the interpretations of the results in terms of migration or demographic changes can clearly not explain all the variation observed. Hence, and in line with the observations of the British Census administration, the only remaining explanation is caste identity manipulation.

---

<sup>24</sup>The data also exists for the year 1921, but is not reported here, as the age categories do not allow to distinguish the age groups born after 1901 from those born before as cleanly as the 1911 data allows. Such information was unfortunately not present in previous Census years.

<sup>25</sup>This amount is found using this calculation:  $Population_{CF39+} = Population_{39-} * 22.27\% / (1 - 22.27\%)$

## 5 Heterogeneity of the effect: access to the canal colonies

We have seen that one of the main advantages given by the agricultural caste status was that it was compulsory to be a member of those castes in order to have access to the land of the canal colonies. One of the specificities of the canal colonies is that they were built in almost desert areas<sup>26</sup>. Hence, the grantees had to come from other regions, making migration play a big role in the development of the canal colonies<sup>27</sup>. Indeed, the Canal colonies have seen their population vastly increase between 1881 and 1921: the population of the districts in which they are located jumped from 5 million to 7.9 million (+57%), while the rest of Punjab remained relatively stable (+7%). This points to a vast migration movement within British districts towards the Canal Colonies<sup>28</sup>. An interesting feature of the process of the colonization of this area of Punjab is that not only did the Punjab Government chose the recipients of the land grants with respect to their caste identity from 1901 on, but they also chose the districts of origin of the “grantees” from the beginning of the colonization scheme. Indeed, among the objectives of the colonization was to “*provide relief from population congestion...*” (Ali, 1988). Hence, only certain districts had access to the canal colonies. Figure 7 presents the districts eligible to the canal colonies according to Ali (1988). Being a member of an agricultural caste thus granted different benefits depending on the district of residence, with the agricultural caste status granting a much larger economic advantage in the districts eligible to the canal colonies land.

[Figure 7 about here.]

This suggests that the incentive to manipulate one’s caste identity was different across districts, and that the districts eligible should exhibit a larger tendency to caste identity manipulation. This calls for a specification separating the eligible districts from the others:

$$\begin{aligned} \ln(pop_{idt}) - \ln(pop_{idt-1}) = & constant + \beta agr_i + \gamma post1901_t + \lambda access_d + \delta agr_i * post1901_t \\ & + \rho access_d * agr_i + \pi agr_i * post1901_t * access_d + \eta post1901 * access_{dt} + \epsilon_{itd} \end{aligned} \quad (5)$$

<sup>26</sup> “These areas [...] were practically desert waste supporting no settled population” (Paustian, 1930).

<sup>27</sup> “According to the Chenab Colony’s final colonization report, the population of the area grew from 112,000 in 1891 to over 1.1 million in 1911, of which the majority were migrants from other parts of the Punjab.” Gilmartin (2004).

<sup>28</sup> “...the Punjab witnessed a major migration from Central Punjab into the newly opened canal colonies of Western Punjab” (Gilmartin, 2004)

With the same notation as in Model 1 and  $access_d$  a dummy indicating whether district  $d$  had access to the canal colonies, or was itself a canal colony. Table 6 shows the results of the regression. The first column at the “Access level” aggregates the population of each type of British districts, those that have no access to the canal colonies, and those that have access to them/are a canal colony, in effect splitting the British Punjab in two parts. It can be seen that while the movement of caste identity manipulation is widespread throughout the districts of Punjab, as indicates the coefficient on  $post1901 * agricultural$  (even though it loses significance when district dummies, year dummies and their interactions are added as control), it is much more important in the districts that have access to the canal colonies, as the positive and significant coefficient on  $post1901 * access * agricultural$  indicates<sup>29</sup>.

[Table 6 about here.]

It is thus shown that caste identity change was up to three times as large in the districts who had access to the land distributed in the canal colonies, and that this evolution is not driven by migration.

## 6 Mistargeting of land in the canal colonies

### 6.1 Data sources on land ownership in colonial Punjab

This section proposes to focus on the land distribution made by the Punjab government in the Canal colonies. The legislation on the land distribution by the Punjab government has its own history, and is detailed in Online Appendix 7. Interested readers can also refer to Ali (1988), which is the main historical reference of this section.

In order to document the mistargeting of government land due to caste identity manipulation, we have collected land ownership data in colonial Punjab from two main sources: the annual Reports on the Land Revenue Administration of the Punjab (yearly published from 1862 to 1940) and the Reports on the working of the Punjab Alienation of Land Act (yearly published from 1902 to 1909). They contain information on annual sales at the district level (area sold). With the creation of the “agricultural” category, they distinguish between agricultural and members of other castes from 1902 on. In particular, they distinguish the sales between and within the two types of castes for the

---

<sup>29</sup>This result could be an artifact due to a within British district migration towards the canal colonies in two steps: in a first step, a member of an agricultural caste would move from a district not having access to the canal colonies to one such district, and in a second step, would migrate to the canal colonies. Online Appendix 5 rules out this interpretation.

entire period under scrutiny. From 1910 on, they also contain data on land ownership (area owned and number of owners), and distinguish three types of land owners: member of agricultural castes, member of other castes and “shamilat, village abadi and Government property”, which are various forms of communal ownership, the bulk of which is constituted by Government property<sup>30</sup>. In order to acquire the ownership of their land, “colonists” had to have occupied the land allocated to them by the government for at least 10 years<sup>31</sup>. Hence, land ownership data for the period 1911-1931 is relevant for the population that manipulated its caste identity during the 1901-1921 period.

This data thus allows to detail very precisely how the land ownership of agricultural castes has evolved (in terms of number of owners and total area owned), and what has been the origin of this evolution, between the sales between caste groups and the distribution of government’s land in the “canal colonies”. As is described in Online Appendix 12, the total amount of sales between members of agricultural castes and non agricultural castes amounted for a tiny fraction of the land owned<sup>32</sup>, and can not account for the evolution of the land ownership of the agricultural castes. The remaining of this section will thus focus only on the transfers of land from the government to the agricultural castes, having removed from the data all land sales between caste groups<sup>33</sup>.

## 6.2 An estimate of mistargeting of land distribution

The data being aggregated at the district level, it is not possible to link an individual having manipulated her caste identity to the access to government land. We will thus resort to an accounting exercise aiming at bounding the total mistargeting. Hence, this section should be taken as illustrative evidence and the estimates proposed interpreted with much more caution than the rest of the paper.

The method proposed in section 3.3 applied to the coefficients of Table 6 can be used to estimate the number of caste identity switchers in the canal colonies districts and in the districts having access to them. Table 6 gives descriptive evidence of two types of identity manipulation: the coefficient on `post1901*agricultural` would give estimates on individuals that manipulate their identity to get the “standard” policy advantages offered

---

<sup>30</sup>Online Appendix 8 details the construction of the data.

<sup>31</sup>Please refer to Online Appendix 7 for the evolutions of the rules of land ownership acquisition for “colonists”. For more details, Ali (1988) proposes a thorough analysis of the history of the canal colonies.

<sup>32</sup>The total area sold from non agricultural to agricultural castes between 1902 and 1931 represented only 1.3% of the land privately owned in 1931.

<sup>33</sup>For a detailed account of the data construction, please refer to the Online Appendices 8, 9, 11, 13 and 14

to agricultural castes, while the coefficient on `post1901*access*agricultural` would give the number of manipulators aiming at the access to government land. Only the latter are of interest here<sup>34</sup>. However, the coefficient on `post1901*access*agricultural` presented in table 6 is a simple correlation, and is given here a causal interpretation for the sake of this accounting exercise, but one should have this caveat in mind. With the information on the evolution of the number of land owners, we can compute the share of manipulators among the agricultural caste members obtaining land grants from the government and from that, the share of land attributed to “manipulator”. Online Appendix 15 details the computation of the share of manipulators among the beneficiaries of the grant, estimated to be between 14.7% and 32.5% for 1911-1921 and 15% and 31% for 1921-1931. Figure 8 panel (a) pictures the amount of land obtained by caste identity manipulators every year between 1911 and 1931, while panel (b) shows the accumulation of land obtained by manipulators on the total amount of land distributed. Hence, in 1931, the total land obtained by manipulators amounted to between 3.8 and 8.2% of the total area allocated by the government during that period.

[Figure 8 about here.]

## 7 Conclusion

Using various identification strategies, this paper shows that the enactment of the Punjab Alienation of Land Act in 1901, by creating an “agricultural castes” category with almost exclusive access to the land market (a huge economic advantage in a Province of Punjab whose population was still rural at almost 90% in 1921) has created a large movement of caste identity manipulation. Indeed, caste groups were given a very strong incentive to manipulate their caste identity in order to benefit from the Act, and from 1901 on, the trend of the population of agricultural castes exhibited a relative increase of 8 to 12 percentage points per decade depending on the specifications, as compared to the trend of the population of non agricultural castes. As this effect only takes place in the British districts of Punjab and does not vary with the exposure to the various epidemics of the period, we can rule out that the various demographic shocks of the period drive the results. Moreover, we show that neither migration nor demography can explain this evolution, underlining that the results are mainly driven by the ability of caste groups to

---

<sup>34</sup>See Online Appendix 6.2 for the detail of the calculation. It is to be noted that since it is only the total amount of caste identity manipulators that need to be equal between non agricultural caste and agricultural castes, this method proposes two estimates for the number of caste identity manipulators that manipulated *in response to the access to government land*.

manipulate their identity in response to administrative incentives, and that up to 3.8% of the total population (7.5% of the agricultural castes population) manipulated its caste identity in order to benefit from the protection of the Punjab Alienation of Land Act. It is estimated that this identity manipulation led to a mis allocation of government land of up to 8.2%

## References

- Akerlof, George and Rachel Kranton**, “Economics and Identity,” *Quarterly Journal of Economics*, 2000, 114 (3), 715–753.
- Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin Olken, and Julia Tobias**, “Targeting the Poor: Evidence from a Field Experiment in Indonesia,” *American Economic Review*, Forthcoming.
- Ali, Imran**, *The Punjab under Imperialism, 1885-1947*, Oxford University Press, 1988.
- Assayag, Jackie**, “The Making of Democratic Inequality. Caste, Class, Lobbies and Politics in Contemporary India (1880-1995),” *Pondy Papers in Social Sciences*, 1995.
- Athreya, Siva and Rohini Somanathan**, “Quantifying Spatial Misallocation in Centrally Provided Public Goods,” *Economic Letters*, 2008, 98 (2), 201–206.
- Ban, Radu and Vijayendra Rao**, “The Political Construction of Caste in South India,” *Working Paper*, 2007.
- Barrier, Norman G**, *The Punjab Alienation of Land Bill of 1900*, Duke University, 1966.
- Bayly, Susan**, *Caste, Society and Politics in India*, Cambridge University Press, 1999.
- Bodenhorn, Howard and Christopher Ruebeck**, “The Economics of Identity and the Endogeneity of Race,” *NBER working paper 9962*, 2003.
- Botticini, Maristella and Zvi Eckstein**, “From Farmers to Merchants, Conversions, and Diaspora: Human Capital and Jewish History,” *Journal of the European Economic Association*, 2007, 5 (5), 885–926.
- Camacho, Adriana and Emily Conover**, “Manipulation of Social Program Eligibility,” *American Economic Journal: Economic Policy*, May 2011, 3 (2), 41–65.



- Cameron, Colin, Jonah Gelbach, and Douglass Miller**, “Robust Inference with Multi-way Clustering,” *Journal of Business and Economic Statistics*, August 2011, 29 (2), 238–249.
- Caselli, Francesco and Wilbur John Coleman**, “On the Theory of Ethnic Conflict,” *Working Paper*, 2011.
- Conlon, Frank F.**, “The Census of India as a Source for Historical Study of Religion and Caste,” in Gerald Barrier, ed., *The Census in British India: New Perspectives*, Manohar, 1981.
- Dirks, Nicholas B.**, *Castes of Mind. Colonialism and the Making of Modern India*, Princeton University Press, 2001.
- Dudley-Jenkins, Laura**, *Identity and Identification in India. Defining the Disadvantaged*, RoutledgeCurzon, 2003.
- Duflo, Esther and Sendhil Mullainathan**, “How Much Should We Trust Differences-in-Differences Estimates ?,” *Quarterly Journal of Economics*, February 2004, 119 (1), 249–275.
- Eynder, Oliver Vanden**, “Military Service and Human Capital Accumulation: Evidence from Colonial Punjab,” *Working Paper*, 2011.
- Francis, Andrew and Maria Tannuri-Pianto**, “Endogenous Race in Brazil: Affirmative Action and the Construction of Racial Identity Among Young Adults,” *Economic Development and Cultural Change*, 2013.
- Ghurye, Govinda**, *Caste and Race in India*, London: Routledge and Kegan Paul, 1932.
- Gilmartin, David**, *Empire and Islam. Punjab and the Making of Pakistan.*, University of California Press, 1988.
- , “Migration and Modernity: the State, the Punjabi Village and the Settling of the Canal Colonies,” in Ian Talbot and Shinder Thandi, eds., *People on the Move. Punjabi Colonial and Post Colonial Migration.*, Oxford University Press, 2004.
- Guha, Sumit**, “The Politics of Identity and Enumeration in India c. 1600-1990,” *Comparative Studies in Society and History*, 2003, 45 (1), 148–167.

- Islam, M. Mufakharul**, “The Punjab Alienation Act and the Professional Moneylenders.,” *Modern Asian Studies*, 1995, 29, 271–291.
- Iyer, Lakshmi**, “Direct versus Indirect Colonial Rule in India: Long-term Consequences.,” *The Review of Economics and Statistics*, 2010, 92 (4).
- Kaul, Pandit Harikishan**, *Report on the Census of Punjab 1911 1912*.
- Khan, Khan Ahmad Hasan**, *Report on the Census of Punjab 1931 1933*.
- Lal, Shadi**, *Commentaries on the Punjab Alienation of Land Act*, Lahore University Book Agency, 1937.
- Liebertson, Stanley and Mary C. Waters**, “The Ethnic Responses of Whites: What Causes their Instability, Simplification, and Inconsistency ?,” *Social Forces*, December 1993, 72 (2), 421–450.
- Michalopoulos, Stelios**, “The Origins of Ethnolinguistic Diversity: Theory and Evidence,” *American Economic Review*, 2012, 102 (4).
- Nagel, Joane**, “American Indian Ethnic Renewal: Politics and the Resurgence of Identity,” *American Sociological Review*, 1995, 60, 947–965.
- Paustian, Paul**, “Canal Irrigation in the Punjab: an Economic Inquiry Relating to Certain Aspects of the Development of Canal Irrigation by the British in the Punjab.” PhD dissertation, Columbia University 1930.
- Punjab Government, ed.**, *Annual Report on the Punjab Alienation of Land Act, for the year ending the 30th September 1905*, Lahore Civil and Military Gazette Press, 1906.
- , **ed.**, *Annual Report on the Punjab Alienation of Land Act, for the year ending the 30th September 1907*, Lahore Civil and Military Gazette Press, 1908.
- , **ed.**, *Annual Report on the Punjab Alienation of Land Act, for the year ending the 30th September 1908*, Lahore Civil and Military Gazette Press, 1909.
- Risley, Herbert**, *Report on the Census of India, 1901 1903*.
- Rose, H.A.**, *Glossary of the Tribes and Castes of the Punjab and North West Frontier*, University of California Press, 1911.

- Rudolph, Lloyd I. and Susanne Hoerber Rudolph**, “The Political Role of India’s Caste Associations,” *Pacific Affairs*, 1960, 33 (1), 5–22.
- Saez, Emmanuel**, “Do Taxpayers Bunch at Kink Points ?,” *American Economic Journal: Economic Policy*, 2010, 2 (3).
- Somanathan, Rohini**, “La Demande pour le Désavantage,” *Afrique Contemporaine*, 2008, 2 (226), 111–131.
- Srinivas**, *Social Change in Modern India*, University of California Press, 1966.
- Talbot, Ian**, “The Punjab under Colonialism: Order and Transformation in British India,” *Journal of Punjab Studies*, 2007, 14 (1), 3–10.
- The Hindu**, “Fake Caste Certificates Racket for Getting Admission to DU Detected,” *The Hindu*, June, 27 2011.
- Yong, Tan Tai**, *The Garrison State: Military, Government and Society in Colonial Punjab, 1849-1947*, Sage Publications, 2005.

## Appendices

### A Disease environment robustness check

This appendix complements section 3.2 by proposing an additional robustness check on the impact of the disease environment on the relative evolution of the two caste groups. The Reports on the Sanitary Administration of Punjab contain the yearly deathrates of each British district<sup>35</sup>. We can then control by the extent to which each district was affected by the disease, and see if it is the districts that were the most affected that saw their agricultural castes have their share in the population increase the most. We then create the variable “deathrate” which center and reduces the average deathrate of each district over each decade, allowing for an interpretation of the coefficient on “deathrate” as the effect of an increase by a standard deviation of the death rates. Table 7 shows the results of the regression of the growth rates of caste groups on their agricultural status interacted with the difference with the average death rate. It can be seen that the coefficient on *post1901 \* agricultural* is not affected, while the coefficient

---

<sup>35</sup>We thank Dave Donaldson for having shared this data.

on  $post1901 * agricultural * deathrate$  is not significant, pointing to the fact that the districts more affected by the epidemics do not exhibit a significantly different pattern than the average district.

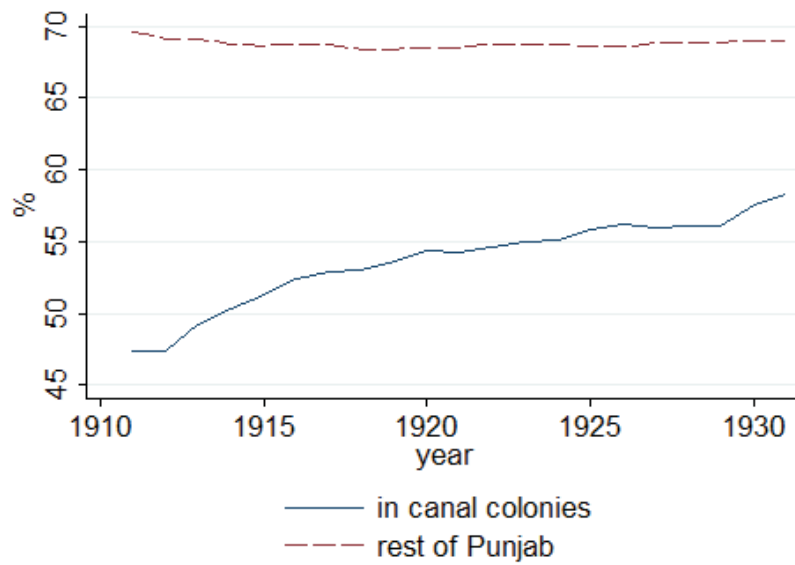
[Table 7 about here.]

## B Placebo tests

An other test for the causal impact of the Punjab Alienation of Land Act on the caste composition of Punjab is to resort to placebo tests: it might be the case that the difference in the growth rates of the population of the two caste groups often switches sign, and that it so happens that such a change happened around 1901. To test for this eventuality, we will resort to simple placebo tests, showing that the only time at which a significant change in the difference of the growth rates between agricultural castes and non agricultural castes happened was around 1901. we will reproduce the same regression as described in Model 1, but using three different time windows: 1881-1901, with the turning point put at 1891, 1891-1911, with the turning point put at 1901 and finally 1901-1921 with the turning point at 1911. If it is really the Alienation Act causing the change in the difference in the growth of the two caste groups, then only the coefficient on  $post*agricultural$  associated to the 1901 turning point should be positive and significant, while the two other turning points should have a small and non significant coefficient. Figure 9 pictures the three coefficients on  $post*agricultural$ , with a varying turning point. It can be seen that the only coefficient positive and significant is the one associated with 1901, in line with the Alienation of Land Act affecting the caste composition of Punjab.

[Figure 9 about here.]

Figure 1: Evolution of agricultural castes' land ownership, by Canal colony status.



Source: Reports on the Land Revenue Administration of Punjab, 1911-1931

Figure 2: British Punjab : dropped and merged districts

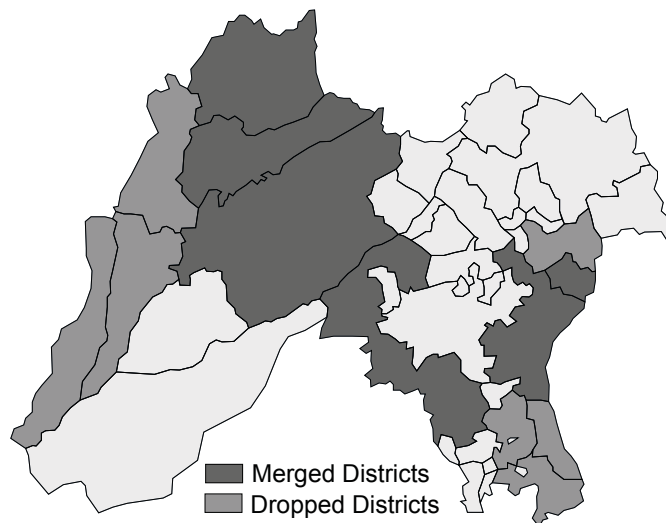


Figure 3: British Punjab: Princely States and British Districts

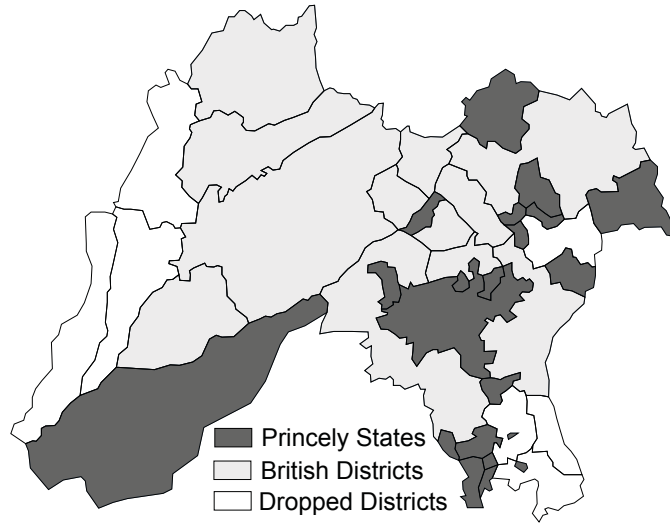
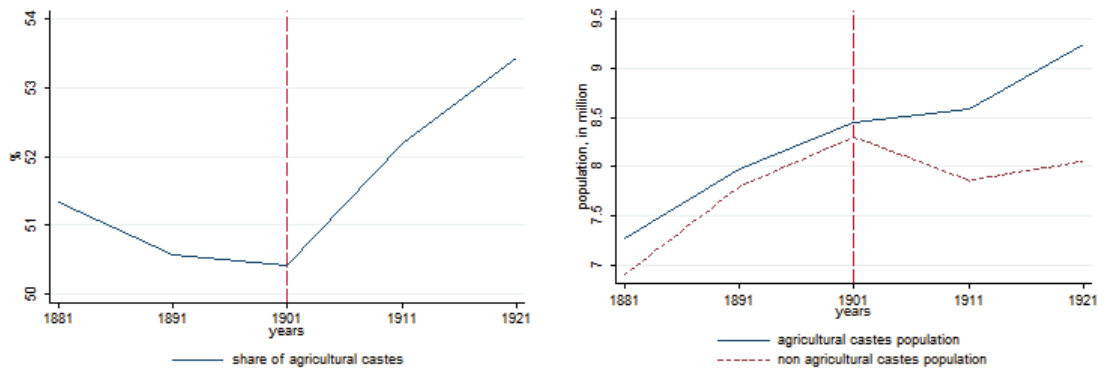
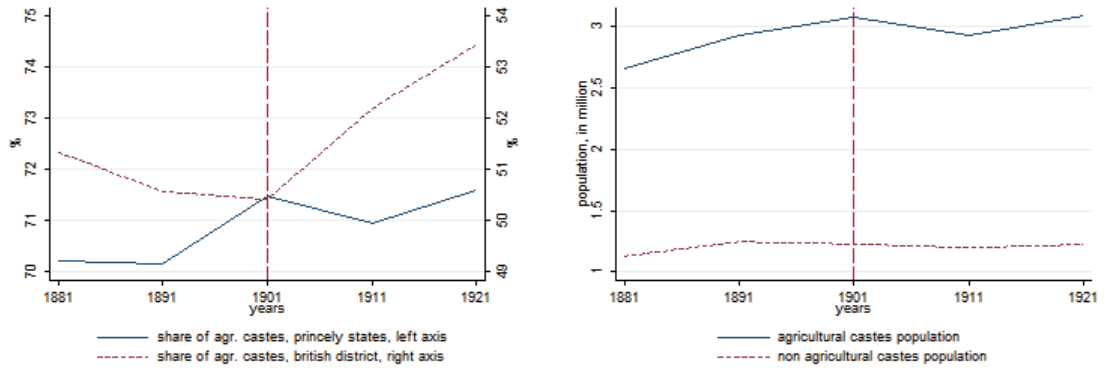


Figure 4: Evolution of the populations of agricultural versus non agricultural tribes in British districts of Punjab, 1881-1921.



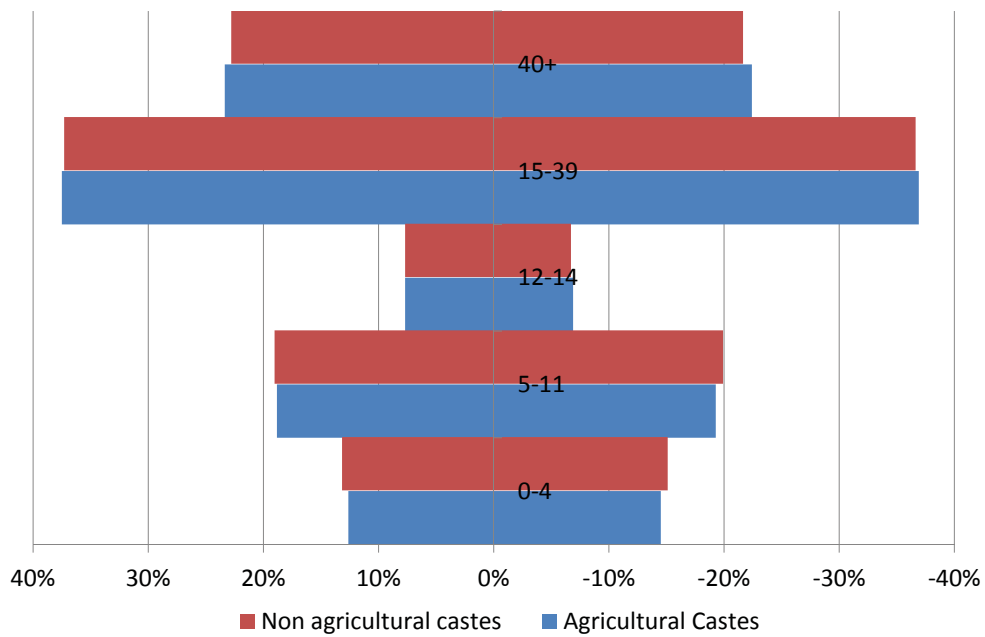
Source: Reports on the Census of Punjab, 1881 to 1921.

Figure 5: Evolution of the populations of agricultural versus non agricultural tribes in the Princely States of Punjab. 1881-1921.



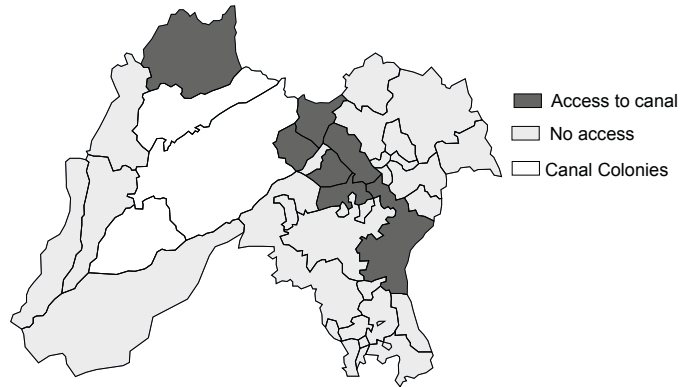
Source: Reports on the Census of Punjab, 1881 to 1921.

Figure 6: Age pyramid by agricultural status, 1911.



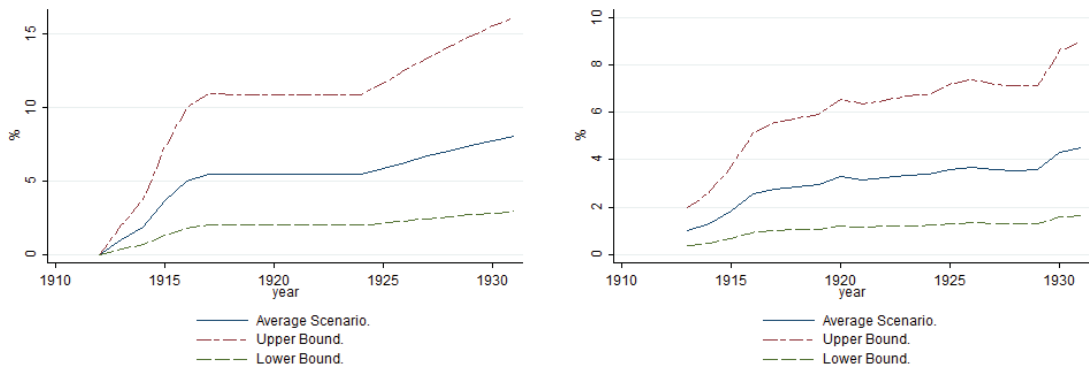
Source: Report on the Census of Punjab, 1911.

Figure 7: Districts whose population is eligible to land in the Canal colonies.



Source: Ali (1988)

Figure 8: Land obtained by caste identity manipulators.



(a) Share of the area distributed obtained by manipulators each year.

(b) Cumulative share of land obtained by manipulators.



Figure 9: Placebo tests regressions.

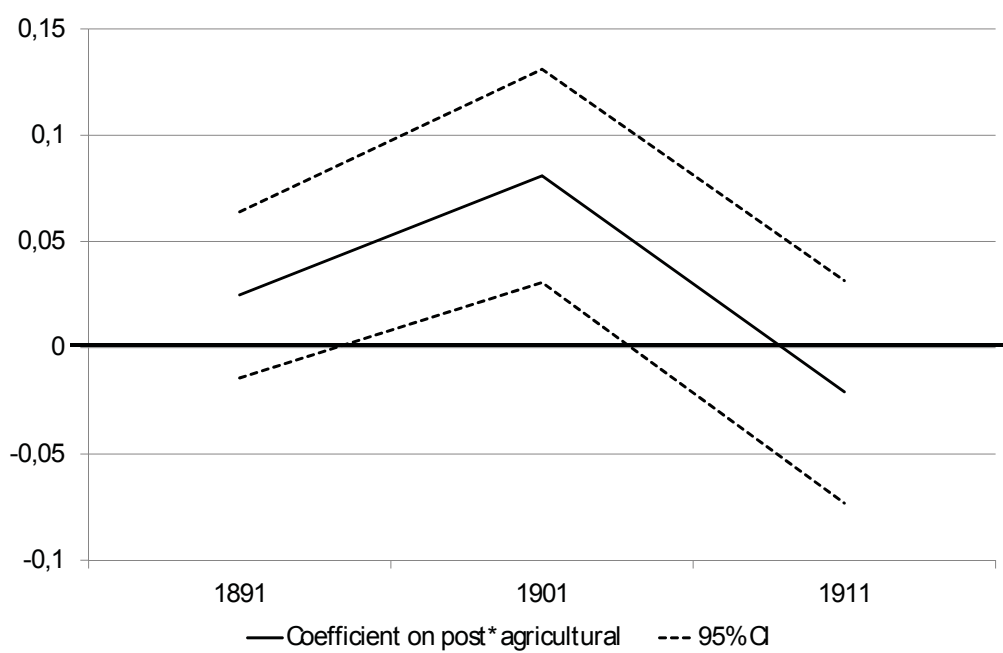


Table 1: Descriptive Statistics: districts and states of Punjab, 1901.

	British Districts	Princely States
Mean Population (std error)	1,408,241 (1,081,661)	207,298 (357,096)
Mean Population/km2 (std error)	291 (175)	194 (127)
Mean Urban Population (std error)	10.8% (0.05)	9.9%(0.08)
Number of Districts/States	12	21

*Source:* Report on the Census of Punjab, 1901. The figures refer to the districts made comparable over time.

Table 2: Descriptive Statistics: population of castes by agricultural status, 1901.

	Agricultural Castes	Non agricultural castes
Mean population (std deviation)	506,789 (930,499)	86,496 (200,510)
N	26	91

*Source:* Report on the Census of Punjab, 1901.

Table 3: Impact of the Alienation Act.

	Province level	District level	
	(1)	(2)	(3)
<b>post1901*agr</b>	<b>0.0799***</b>	<b>0.0789***</b>	<b>0.0751***</b>
	(0.0186)	(0.0122)	(0.0222)
agr	-0.0177	-0.0174***	-0.00564
	(0.0126)	(0.00571)	(0.0146)
post1901	-0.109***	-0.116***	
	(0.0136)	(0.0269)	
Decade Dummies	NO	NO	YES
District Dummies	NO	NO	YES
Decade*District Dummies	NO	NO	YES
Observations	429	2640	2640
Adjusted $R^2$	0.210	0.068	0.264

Weighted OLS regressions of caste's population growth rates by decade. Standard errors are clustered at the caste level in column 1 and are two way clustered at the district-decade and caste levels in column 2 and 3, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Table 4: British districts vs Princely States.

	Province level	District level	
	(1)	(2)	(3)
<b>post1901*british*agr</b>	<b>0.102***</b> (0.0329)	<b>0.127**</b> (0.0606)	<b>0.126*</b> (0.0702)
<b>post1901*agr</b>	<b>-0.0222</b> (0.0207)	<b>-0.0488</b> (0.0542)	<b>-0.0517</b> (0.0606)
agr	0.0338*** (0.0123)	0.0434 (0.0427)	0.0440 (0.0472)
british*agr	-0.0514*** (0.0157)	-0.0595 (0.0432)	-0.0486 (0.0492)
post1901*british	-0.0612** (0.0255)	-0.0915 (0.0654)	
post1901	-0.0477** (0.0183)	-0.0233 (0.0560)	
british	0.0528*** (0.00980)	0.0610* (0.0347)	
Decade Dummies	NO	NO	YES
District Dummies	NO	NO	YES
Decade*District Dummies	NO	NO	YES
Observations	753	5725	5725
Adjusted $R^2$	0.168	0.051	0.222

Weighted OLS regressions of caste's population growth rates by decade. Standard errors are clustered at the caste level in column 1 and are two way clustered at the district-decade and caste levels in column 2 and 3, \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 5: Migration robustness check.

	Province level	District level	
	(1)	(2)	(3)
<b>post1901*agr</b>	<b>0.0545***</b> (0.0186)	<b>0.0516***</b> (0.0152)	<b>0.0474*</b> (0.0264)
agr	-0.0177 (0.0126)	-0.0162** (0.00631)	-0.00452 (0.0156)
post1901	-0.0949*** (0.0136)	-0.101*** (0.0259)	
Decade Dummies	NO	NO	YES
District Dummies	NO	NO	YES
Decade*District Dummies	NO	NO	YES
Observations	429	2850	2850
Adjusted $R^2$	0.169	0.052	0.243

Weighted OLS regressions of caste's population growth rates by decade. Standard errors are clustered at the caste level in column 1 and are two way clustered at the district-decade and caste levels in column 2 and 3, \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 6: Heterogeneity of the effect: access to the canal colonies

	Access level		District level	
	(1)	(2)	(3)	
<b>post1901*access*agr</b>	<b>0.0466**</b> (0.0177)	<b>0.0460***</b> (0.00821)	<b>0.0528**</b> (0.0249)	
<b>post1901*agr</b>	<b>0.0389***</b> (0.0132)	<b>0.0370***</b> (0.00857)	<b>0.0289</b> (0.0189)	
agr	-0.0297*** (0.00911)	-0.0279*** (0.00599)	-0.00989 (0.0163)	
access*agr	0.0147 (0.0109)	0.0136** (0.00540)	0.00625 (0.0255)	
post1901	-0.0694*** (0.00735)	-0.0672* (0.0344)		
access	-0.00426 (0.00704)	-0.00861 (0.0358)		
post1901*access	-0.0452*** (0.0154)	(0.0453)		
Year Dummies	NO	NO	YES	
District Dummies	NO	NO	YES	
Year*District Dummies	NO	NO	YES	
Observations	720	2850	2850	
Adjusted $R^2$	0.190	0.067	0.251	

Weighted OLS regressions of caste's population growth rates by decade. Standard errors are clustered at the caste level in column 1 and are two way clustered at the district-decade and caste levels in column 2 and 3, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Table 7: Controlling for diseases.

	Controlling for death rate	
	(1)	(2)
<b>post1901*agr</b>	<b>0.0629***</b>	<b>0.0573***</b>
	(0.00637)	(0.0215)
<b>post1901*agr*deathrate</b>	<b>0.0204</b>	<b>0.0153</b>
	(0.0268)	(0.0318)
agr*deathrate	-0.0207	-0.0251
	(0.0282)	(0.0316)
agr	-0.00441	0.0126
	(0.00966)	(0.0176)
post1901	-0.0898***	
	(0.0346)	
deathrate	-0.00409	
	(0.0244)	
post1901*deathrate	-0.0351	
	(0.0324)	
Decade Dummies	NO	YES
District Dummies	NO	YES
Decade*District Dummies	NO	YES
Observations	2079	2079
Adjusted $R^2$	0.075	0.274

Weighted OLS regressions of caste's population growth rates by decade. Standard errors are two way clustered at the district-decade and caste levels, \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .