Caste dominance and economic performance in rural India[+]

Vegard Iversen[£], Adriaan Kalwij[$], Arjan Verschoor* and Amaresh Dubey**

[£]IDPM, University of Manchester, e-mail: Vegard.Iversen@manchester.ac.uk

[$] Utrecht University, *University of East Anglia, * *Jawaharlal Nehru University

ABSTRACT

Using household panel data for rural India covering the years 1993/94 and 2004/05, we test whether Scheduled Caste (SC) and other minority groups perform better or worse in terms of income when resident in villages dominated by (i) upper castes or (ii) their own group. Theoretically, upper caste dominance comprises a potential 'proximity gain' and offsetting, group-specific 'oppression' effects. For SCs and OBCs, initial proximity gains dominate negative oppression effects because upper caste dominated villages are located in more productive areas: once agroecology is controlled for, proximity and oppression effects cancel each other out. Albeit theoretically ambiguous, we find large, positive own dominance or enclave effects for Upper Castes, OBCs and especially SCs. These village regime effects are restricted to the Hindu social groups. Combining pathway and income source analysis, we close in on the mechanisms underpinning identity-based income disparities; while education matters, land ownership accounts for most enclave effects. A strong post reform SC own village advantage turns out to have agricultural rather than non-farm or business origins. We also find upper caste dominance to inhibit the educational progress of other social groups along with negative enclave effects on the educational progress of Muslim women and ST men.

I INTRODUCTION

IA. Aim and motivation

In economics, various mechanisms are recognized that, in a stratified society, link economic welfare with signifiers of social identity such as caste, religion and ethnicity. Some mechanisms originate in 'taste-based' (e.g. Becker 1971) or 'statistical' (e.g. Arrow 1972) discrimination by others and are *external* to the affected group. Other mechanisms are *internal* and hinge on not how a group is seen and treated by others, but how its members perceive and interact among themselves. The consequences for economic performance of a self-image that group membership imparts (Akerlof and Kranton 2000) and of the onset of collective inertia (e.g. Peyton Young 2001) are two examples.

In this paper we explore empirically the proposition that the balance of forces linking social identity to economic performance is influenced by the relative economic or political power of the various social groups that live and work in each other's vicinity, to be precise, reside in the same village. We undertake this empirical inquiry for rural India, whose village communities can be seen as a paradigm of social stratification (e.g. Deshpande 2001 and 2011; Anderson 2011).

We study three complementary explanations for identity-based disadvantage. The first, the oppression hypothesis, originates in M. N. Srinivas's theory of caste dominance[1] which portrays a caste that apart from strong numerical presence is also economically powerful (Srinivas 1955). The oppression hypothesis captures the external mechanisms linking social identity and economic welfare and suggests that historically disadvantaged and other marginalised social groups fare worse when resident in villages dominated by upper castes.

The second, the village enclave hypothesis, corresponds with the internal mechanisms linking identity and welfare, is theoretically ambiguous and depicts a situation where a marginalised group is dominant at the village level. Upwards mobility may then be inhibited, or conversely encouraged, by factors internal to the group in question. To illustrate, the absence of role models or a preference for traditional occupations could lock individuals of marginalised backgrounds into low level equilibrium traps (Akerlof and Kranton 2000). By reducing the social distance between parties to rural transactions, own enclaves could also improve the operation of vital rural markets (Anderson 2011).

Thirdly, we evaluate the merit of the proximity hypothesis, which is anchored in a theory of public goods provision and suggests that minority groups may benefit from being proximate to politically well-connected and prosperous upper castes (e.g. Sethi and Somanathan 2010). We explain why proximity and oppression provide complementary insights about the roots of caste-based disparities in rural India.

Our paper adds a timely political economy dimension and new empirical insights to the literature addressing identity, economic disadvantage and its persistence. Existing studies linking economic performance to the village level balance of power are few and Anderson (2011) is the only other comprehensive effort.

Pertaining to India and in spite of bold legislation that made reservations of government jobs and seats in legislative assemblies and educational institutions a hallmark policy, households of Scheduled Caste[2] (former 'untouchables') and Scheduled (indigenous) Tribe backgrounds continue to feature disproportionately on key indicators of rural deprivation.[3] This persistence remains a puzzle that we attempt to shed new light on.

Finally, India's so-called 'silent revolution' manifested in the rapid rise in lower caste representation in state-level legislative assemblies (Jaffrelot 2003), suggests that a key ingredient for social change already is in place. Banerjee and Somanathan's (2007) study of parliamentary constituencies and rural infrastructure provision between 1971 and 1991 supports this view since social groups that politically mobilised, namely Scheduled Castes, appear to have leaped forward relative to those that did not (Scheduled Tribes and Muslims).

We see two reasons for questioning the growing optimism about the remedial and transformative potential of the democratic process whether on its own or aided by political reservations (e.g. Pande 2003). Firstly, the data used in previous studies are too coarse to undertake the necessary welfare and poverty comparisons: village variables do not adequately account for (infrastructure) quality variation while state level expenditure and other variables do not capture benefit incidence and the magnitudes of improvements in enough depth. The second is the analytical bypass of village level institutional hurdles to social change. With the emergence of a new, rich dataset described in detail below, we aim to remedy this neglect.

IB. Background and contribution to the literature

'March 1949: A group of Scheduled Caste members from villages around Delhi had been thrown out of their homes by Jat landowners angered that these previously bonded servants had the cheek to take part in local elections and graze their cattle on the village commons.

June 1951: A village in Himachal Pradesh. A conference of Scheduled Castes is attacked by Rajput landlords. The SCs are beaten up with sticks, their leaders tied up with ropes and confined to a cattle pound.

June 1952: A village in the Madurai district of Madras State. A SC youth asks for tea in a glass at a local shop. Tradition entitles him only to a disposable coconut shell. When he persists, he is kicked and hit on the head by caste Hindus.

June 1957: A village in the Parbani district of Madhya Bharat. Newly converted Buddhists [previously "untouchable" Hindus] refuse to flay carcasses of dead cattle. They are boycotted by the Hindu landlords, denied other work and threatened with physical reprisals.' (Guha 2007, 380-81)

More than 50 years later and in spite of a weakening of the more forbidding caste barriers[4], Scheduled Caste (SC) and Scheduled Tribe (ST) households remain overrepresented among India's rural poor, illiterate and in the former case, also the landless. While rural poverty is declining, these two groups, which represent 16.2 (SC) and 8.2 (ST) percent of the country's population,[5] account for 47.3 percent of India's rural poor (Gang et al. 2008a). A less sharply delineated category of disadvantaged citizens mentioned by the Constitution, Other Backward Classes (OBC), also continues to have lower living standards than the mainstream population (Gang et al. 2008b).[6] The results reported below suggest that the same holds for

Muslims, the largest religious minority accounting for 13.4 percent of the population (Census of India 2001).

Shah et al.'s (2006) study of untouchability, covering 550 villages in 11 main states, found that SCs were prevented from full participation in local markets and often from entering village shops in 30-40 percent of the villages surveyed; in 45-50 percent of these villages, SCs were prevented from selling milk to village dairy cooperatives. Such 'bans' are rooted in purity and pollution ideals and the ensuing and sensitive links between a person's caste and the preparation and handling of food and water (e.g. Madsen 1991, Iversen and Raghavendra 2006). Indeed, and well known, SC hamlets tend to be separate from the main village and often have their own drinking water source.[7]

We test our hypotheses by examining the relationship between the social identity of the groups that are economically or numerically dominant at the village level and the income of households belonging to marginalised groups advancing the literature as follows. Firstly, a few studies test for identity-based disadvantage in India (e.g. Kijima 2006; Gang et al. 2008a), but do not test whether village level upper caste or own group dominance affect economic performance.[8] In addition, little remains known about whether and in what directions, patterns, magnitudes and causes of identity-based disadvantage have transmuted during the post reform years.

Secondly, we broaden the remit of empirical research on identity aspects of economic performance. In India, empirical research on caste has focused mainly on labour market discrimination (e.g. Banerjee and Knight 1985; Kingdon 1998; Thorat and Attewell 2007). Evidence suggests that individuals of SC and ST background are indeed disadvantaged – through lower wages, a higher propensity of being stuck in

dead end jobs (e.g. Banerjee and Knight 1985) or inferior employment terms, such as casual employment (e.g. Madeshwaran and Attewell 2007).[9]

To date, much anecdotal but little systematic knowledge exists about discrimination in credit, insurance or other key markets or particular to rural areas, markets for agricultural inputs and outputs. There is also limited evidence on whether caste, religious or tribal identity circumscribes the access to poverty-oriented public policy programmes or public services in general.[10]

Thirdly, we provide a major push forward of the empirical literature using sociological and anthropological notions of caste dominance where Anderson (2011) is the other main contributor. For a data-set covering 120 villages in Uttar Pradesh and Bihar, she observes that Yadav households in villages where Yadavs are the dominant land owners have higher incomes than Yadav households in villages where the dominant land owners belong to a local upper caste. Anderson attributes this result to the market for irrigation water's failure to operate in villages with upper caste land dominance and concludes that social distance may prevent the efficient operation of vital rural markets.

Unlike Anderson (2011), we distinguish first theoretically and then in our empirical specifications, to the extent that these specifications allow, between the potential 'proximity gain' for Scheduled Castes and other social groups from residing in upper caste dominated villages and offsetting, social group specific, oppression effects within the same villages. In the light of Sethi and Somanathan (2010) this distinction is crucial for obtaining a balanced understanding of the origins of caste-based disparities in rural India.

IC. Empirical strategy and main findings

We use a unique household panel data set for rural India to make detailed standard of living comparisons across social groups at two points in time – before the effects of the 1991 liberalisation reforms had started to kick in (1993/94; round 1), [11] and 11 years later (2004/05; round 2). We exploit our access to uniquely detailed information on the largest landowning and population groups in villages where panel households reside to explore three possible, complementary explanations for identity-based disadvantage in rural India.

We identify the effects on household income of belonging to a particular social group, of belonging to a particular social group *and* living in a village dominated by upper castes and ditto but living in a village dominated by one's own group (the 'enclave' effect). The effect of living in an upper caste dominated village represents the net of the 'oppression' and the 'proximity' effect (regardless of which group one belongs to). The proximity effect captures the idea that upper castes are likely to be prosperous, politically well-connected, and able to ensure better access to rural infrastructure and other public goods. [12] In our main regression specification, we first condition on the state of residence and agro-ecology at the district level and proceed to control for contemporaneous household and village level characteristics.

In upper caste dominated villages, we initially find that proximity gains dominate oppression effects for OBCs and SCs: this is not, as others have suggested, because of a better access to village public goods but because such villages are located in more productive areas. Once agroecology is controlled for, this net gain disappears and proximity and oppression effects cancel each other out.

We do, moreover, find large, positive own dominance or enclave effects on income for UCs, OBCs and especially for SCs in the post reform era. A striking

finding is that these village regime effects on income are confined to the Hindu social groups in our panel. We also find upper caste dominance to inhibit the educational progress of other social groups and negative own enclave effects on the educational progress of Muslim women and Schedule Tribe men.

We use the estimated coefficients in our main specification to compute counterfactual income and poverty figures. Upper caste dominance brings an own-group advantage of about 10 percent of mean income in both survey rounds. Even if other groups benefited as much as UCs from a positive externality conferred on them, this proximity gain is more than offset by group specific 'oppression' effects for SC and OBC households, which in round 2 depress mean income of SC and OBC households in upper caste (UC) dominated villages by, respectively, about 14 and 12 percent and raise the percentage in poverty by, respectively, 6 and 5 points. Although the second effect is larger, the net effect of proximity and oppression is, as noted above, statistically insignificant after controlling for agroecology and state of residence.

We confirm robustness of our main results to how dominance is measured: whether as a zero/one variable, which we prefer for parsimony and ease of interpretation, or as the share of village land held by the dominant group, or as a dominance-adjusted Herfindahl index capturing that if land holdings among the non-dominant groups are more fragmented, the intensity of the largest group's dominance should be expected to increase.

Finally, we combine pathway and income source analysis to explore the mechanisms through which these village regime effects manifest themselves. Pathways are explored by gradually introducing sets of variables that capture village infrastructure, household education and household land. It transpires that village

infrastructure has no, education negligible and land the largest such effect: once all three are controlled for, virtually no village regime effects remain. A key insight from the income source analysis is that the resilient SC round 2 enclave effect has agricultural and not, as perhaps expected, non-farm or business roots.

The paper is laid out as follows. Section II describes the data set, elaborates on the theoretical background and presents the empirical model for testing our hypotheses. Section III presents descriptive statistics on income and poverty levels and change and on human capital endowments by social group and village regime. Section IV presents the main empirical results, followed by robustness tests, and a computation of counterfactual income, growth and poverty to illustrate the order of magnitude of the village regime effects that we identify. Section V concludes.

II. DATA, THEORETICAL BACKGROUND AND EMPIRICAL FRAMEWORK

IIA. The data set

The data are from two large-scale household surveys that cover most of the territory of India, the earlier known as the Human Development Profile of India (HDPI) surveys, and the later as the Indian Human Development Survey (IHDS). The first round, HDPI-I (1993/94), was carried out by the National Council of Applied Economic Research (NCAER) on behalf of UNDP. The second round, HDPI-II IHDS (2004/05), was carried out by NCAER on behalf of the University of Maryland. The primary purpose of the surveys was to collect detailed information on a large range of human development indicators, including income, the variable reported on here.

These surveys are the first major ones for India to measure household income in a comprehensive and refined manner, including carefully assessed income from cultivation, self-employment and a large number of other sources (Desai et al. 2009; 16).[13]

The way in which data on income is collected in both rounds is identical or similar for all sources with the exception of crops. At the national level, the figures suggest an annual rural income per capita growth of 4.2 percent, a poverty headcount ratio of 38.3 percent for 1993/94, and of 29.0 percent for 2004/05. These figures are very close to estimates of the incidence and decline of rural poverty in India based on the National Sample Survey Organisation (NSSO) Consumer Expenditure Surveys (CES), despite the use of expenditures, not income in the latter. According to the 50th NSSO CES (1993/94) 37.1 percent of the rural population were in poverty; in the 61st round (2004/05) rural poverty stood at 28.4 percent.[14]

A unique feature of these data is that a village questionnaire was administered in the second round and enables the construction of village social composition and land ownership distribution variables by *jati* (sub-caste). Further, the sub-division of social groups in the household questionnaires allows us to precisely identify the *jati* of individual households and thus to make comparisons of the economic performance of other social groups with that of upper-caste households, who mostly are Hindus.[15] These features depart notably from official data sets with collection of information on *jati* terminated after the 1931 Census.

The first round of the survey used a random sample of households located in and representative of each of the rural areas in all (then sixteen) India's major states. The attrition rate between the two rounds is 18 percent, and due to recontact details not being available in two states,[16] and migration (of the entire household) and natural

demise (Desai *et al.* 2009; 3). After removing about 20 villages with missing social composition and land ownership information, our panel comprises of 9,108 households spread over 679 villages.

Since a residence-based sampling rule was adopted, the findings reported here are strictly speaking valid only for households who choose not to migrate (e.g. Baulch and Hoddinott 2000; Rosenzweig 2003). However, the comparison of living standards and changes therein across social groups – the focus of this paper – should not be much affected by this limitation: the variables caste, religion, education and income are not substantially different in the panel from those in a randomly selected rural refresher sample drawn to check the round 2 representativeness of the panel household sample.[17] Furthermore, we performed a statistical test on whether or not the inclusion in the panel of all households who participated in the first round is associated with our dependent variable household income. After controlling for household demographic composition and educational attainment, household income is not associated with selection into the panel,[18] suggesting no endogenous panel attrition and that our panel households, with respect to income, are a randomly selected subsample of all rural households that participated in the first round.

IIB. Upper caste and own dominance – theory and definitions

The caste dominance concept originates in the sociological and anthropological literature. In Srinivas's (1955, 18) own words:

'*A caste may be said to be 'dominant' when it preponderates numerically over the other castes and when it wields preponderant economic and political power. A large and powerful caste group can more easily be dominant if its position in the local hierarchy is not too low.*'

Upper caste dominance is perhaps best expressed as a combination of secular power and ritual status where the latter reflects the Varna hierarchical order with Brahmins topmost among four broad occupational ranks and with former untouchables (SCs) as a separate category. The dominant social group could be defined as the group (i) which represents a larger share of the village population than any other social group (n_d); (ii) owning more village land than any other social group (l_d) (e.g. Dumont 1970); or (iii) both n_d and l_d (e.g. Srinivas 1955). While not exhaustive, (i)-(iii) are alternative measures of secular power.

Numerical strength could translate into village level political muscle especially after the 73rd Constitutional Amendment's elevation of the status and significance of village Panchayats. However, Anderson (2011) finds no effects of population dominance on economic outcomes. As explained below, our empirical focus on land dominance partly reflects a constraint imposed by de facto village structures in rural India but also exploratory regressions supportive of Anderson's (2011) observations and Dumont's (1970) assertion that dominance is rooted in economic power captured by landownership alone.[19]

Conceptually, let the land of village j, L_j, be distributed over m groups where n_i represents the share of the village land that belongs to social group i. Hence,

$$L_j = \sum_{i=1}^{m} n_i = 1 \qquad (1)$$

Definitions: *A dominant social group has the largest share of the village land of any social group. For members of the dominant social group in village j, village j is own group dominated or an own enclave. If the dominant social group in village j is upper caste, village j is upper caste dominated. Upper caste dominance exemplifies a village regime.*

This forms the conceptual backbone for the main analysis with our preferred dominance measure being sociologically anchored and easy to interpret. This preferred measure neglects the relative size of the dominant group's landholdings, as well as fragmentation or concentration among other social groups within a village. We therefore make use of two alternative dominance measures as robustness checks. The first is the share of village land owned by the dominant group, the second a modified Herfindahl index.

The Herfindahl index of concentration for village *j* may be defined as:

$$H_j = \sum_{i=1}^{m} n_i^2 \qquad \text{where} \quad H_j \in (0,1] \tag{2}$$

Situations where two groups have landholdings of equal size would imply considerable concentration, but not dominance. To equip H_j to capture *dominance*, we introduce the following modification:

$$D_j = n_d^2 - \sum_{i \neq d} n_i^2 \tag{3}$$

where n_d is the land share owned by the dominant group. For a given n_d, the more fragmented is the land ownership of other groups, the higher is D_j. In the example above, the value of D_j will be exactly zero, as it should be.

To construct village level dominance measures, we combine village level information on social structure and land ownership with evidence on the hierarchical status of precisely identified jatis. The village questionnaire administered in round 2 identifies the jati of the numerically dominant social group in each village, the percentage of village land this social group owns along with similar information for the next 4-8 most numerous social groups. Anthropological and other relevant evidence (e.g. Jaffrelot 2003) on the status of different jatis is then invoked to develop a more refined upper caste definition as explained in Appendix 1. Given the general inactivity of rural land markets[20] and that land-dominant groups typically hold a much larger share of village land than any other group,[21] we assume that the village regime is identical in rounds 1 and 2.

IIC. Empirical model

The proximity, oppression and enclave hypotheses refer to the extent to which the income level of households from different social groups is affected by the social identity of the dominant land owners in the village of residence. To test these hypotheses, we model the relative differences in income by social group and village regime, controlling for location and household characteristics, as follows (see Appendix 3 for more details):

$$\begin{aligned}\ln(Y_{ht}) = {} & \alpha_{0t} + \alpha_{1t} SC_h + \alpha_{2t} ST_h + \alpha_{3t} MUS_h + \alpha_{4t} OBC_h \\ & + \beta_{1t} SC_h \times DSC_{v(h)} + \beta_{2t} ST_h \times DST_{v(h)} \\ & + \beta_{3t} MUS_h \times DMUS_{v(h)} + \beta_{4t} OBC_h \times DOBC_{v(h)} \\ & + \beta_{5t} UC_h \times DUC_{v(h)} + \gamma_{1t} SC_h \times DUC_{v(h)} + \gamma_{2t} ST_h \times DUC_{v(h)} \\ & + \gamma_{3t} MUS_h \times DUC_{v(h)} + \gamma_{4t} OBC_h \times DUC_{v(h)}\end{aligned}$$

$$+\pi_t X_{ht} + \theta_h + \eta_{v(h)} + \varepsilon_{ht}, \quad (4)$$

Subscript *h* denotes households, *t* time (t={1993/94,2004/05}) and *v(h)* the village of residence of household *h*. Household real per capita income [22] is denoted by *Y* and the five social groups a household can belong to are denoted by *SC* (Scheduled Castes), *ST* (Scheduled Tribes), *MUS* (Muslims), *OBC* (Other Backward Classes) and *UC* (Upper Caste). These are all dummy variables and take the value 1 if a household belongs to this group and 0 otherwise. The village regime is modeled using the dummy variables *DSC*, *DST*, *DMUS*, *DOBC* and *DUC*, which take the value 1 if this particular social group is land dominant in the village of residence and 0 otherwise.

The last three right hand side terms of equation (4) form the error structure of the model. The first two error terms are, respectively, a random household specific effect, θ_h, that is assumed to be independently distributed across households, and a random village specific effect, $\eta_{v(h)}$, which is assumed to be independently distributed across villages. The third error term, ε_{ht}, is an idiosyncratic error term and is assumed to be independently distributed across households, villages and time. The assumption of a random household specific effect, as opposed to a fixed effect, is required because incorporating a household specific fixed effect would make it impossible to identify proximity, oppression and enclave effects since the village regime is constant over time and panel households live in the same village in both rounds. We estimate equation (4) by Least Squares separately for each round and thus allow all parameters to vary over time. Arbitrary correlation between households within a village is accounted for when calculating the standard errors (e.g. Cameron and Trivedi 2005).

The α-parameters refer to the relative income differences between households of different social groups with *UC* as reference group. For instance, the parameter α_{1t}

(x100) corresponding to the variable *SC*, is interpreted as the percentage difference in income between *SC* and *UC* households living in a village dominated neither by SC nor by UC (*ceteris paribus*).

The enclave hypothesis refers to the β-parameters. For instance, the parameter β_{1t} (x100) corresponding to the variable *SC*x*DSC* represents the percentage difference in income between *SC* households living in a village dominated by their own social group and *SC* households living in a village dominated neither by SC nor by UC. Further, the parameter β_5 (x100) corresponding to the variable *UC*x*DUC* is interpreted as the percentage difference in income between UC households living in a UC dominated village and UC households living in a village not dominated by UC. The net of the proximity and oppression effects is captured by the $\tilde{\beta}$parameters. γ_1, corresponding to the variable *SC*x*DUC*, for instance, is interpreted as the percentage difference in income between SC households living in a UC dominated village and SC households living in a village dominated by neither SC nor UC.

In order to disentangle proximity and oppression effects, we use estimated coefficients to compute counterfactual income as if the externality conferred upon other social groups from living in a UC-dominated village is equal to the UC own enclave effect. That is, we assume that the proximity gain for non-UC households is (at most) equal to the UC enclave effect. In practice it is possible that the externality that causes this proximity effect is smaller, so in doing so we provide an upperbound on the (absolute) oppression effects.[23] See the appendix for details.

Following Anderson (2011), who contends that land holding patterns in village India are historically determined, our village regime variables are assumed to be exogenous determinants of (per capita) household income. To the extent that contemporaneous village-level and household-level characteristics such as village

infrastructure, household land and education are correlated with village regime, it is legitimate to think of these as pathways along which the village regime affects household income. This leaves open the possibility that land holding patterns are historically and jointly determined with land quality. For instance, if UCs, on average, were more successful in the scramble for fertile land, the proximity coefficient could simply pick up that upper caste dominated villages are located in areas with greater agricultural potential.[24] In addition, land reforms, which fell within the jurisdiction of individual states after independence, could have upset the historical land ownership patterns that Anderson's identification strategy relies upon. However, and as Besley and Burgess (2000) document, while state level legislation included introducing land ceilings, redistribution of land has, by and large, been evaded because of loopholes and the absence of political commitment (ibid. 394).[25] The most powerful effects on poverty have instead been observed for reforms strengthening tenurial security (ibid.).[26]

To address these two concerns which may cause a violation of our assumption of exogenous village regime variables, we use Palmer-Jones and Sen's (2003) mapping of agroecological zones onto Indian districts[27] and state dummy variables as additional controls. Both sets of variables are included in the vector of control variables (X) in equation (4). X also includes variables for household demographic composition, education and land holdings, and for village infrastructure (the full variable list is reported in Appendix 2).

As noted, all parameters of equation (4) are allowed to vary with time which makes it possible to investigate changes in enclave effects and obtain clues about changes in proximity and oppression effects between the two rounds and in turn the implications for income growth and for poverty incidence and persistence. As

discussed we explore the robustness of the main results to two alternative measures of dominance and for this purpose we replace the dummy dominance variables (e.g. *DUC*) with the upper caste land share (the first alternative) or the value of the dominance adjusted Herfindahl-index (the second alternative, eq. (3)).

III. DESCRIPTIVE STATISTICS

Anchored in Dumont's (1970) conception of caste dominance, as set out above, our empirical focus is on villages where a particular social group owns the largest proportion of village land. The technical challenge posed by separate identification of land and population dominance is discernible from the diagonal of Table I which shows the strong correlation between population and land dominance: for each social group, if it is population dominant, in over 90 percent of cases, it is also land dominant, and vice versa. Table I shows that Upper Caste dominance is the most common village regime, closely followed by villages dominated by OBCs. The number of SC and Muslim dominated villages is comparatively small.

[Insert table I about here]

Table II reports the distribution of households across village regimes and illustrates the extent to which households are clustered in 'own' dominated villages. Such clustering, which can be read off the bold diagonal, is pronounced for STs, UCs, OBCs and Muslims while the SC population is more dispersed. Relevant to the

oppression hypothesis, table II also shows the presence of households from each social group in UC-dominated villages. 45.9 percent of the households residing in such villages are UCs, 26.4 percent SCs and 21.6 percent OBCs. STs and Muslims between them account for 6.1 percent.

[Insert table II about here]

Next, we present descriptive statistics on village regimes that are pertinent to the proximity, oppression and enclave hypotheses.[28] Figure 1 reports round 1 and round 2 mean household per capita incomes and poverty headcount by social groups for villages with (i) upper caste land dominance, (ii) own group land dominance and (iii) the remaining 'other' villages. Unsurprisingly, in the aggregate, SCs and STs are on average worse off than OBCs and Muslims, who are in turn poorer than UCs, which is true in both rounds and whether measured by income or poverty incidence. However, a more nuanced picture is obtained once we compare living standards by social group across village regimes.

[Insert figure 1 about here]

Figure 1 suggests pronounced village regime effects on income levels, growth, poverty incidence and the speed of poverty reduction (or conversely, poverty persistence). In round 1, SCs and OBCs in upper caste dominated villages have marginally higher average incomes. For STs, round 1 incomes outside own enclaves were notably higher. The average upper caste household was much better off in own

enclaves, while Muslim incomes show little variation across village regimes. In terms of how income by social group ranks across village regimes, the second round picture is broadly similar to that of the first round for OBCs, Muslims and UCs but strikingly different for STs and SCs: STs in round 2 appear to do much better in UC dominated villages while SCs fared much better in own enclaves.

In terms of average living standard improvements, enclaves seem to favour UCs and SCs very strongly and Muslims marginally; STs did remarkably well in UC dominated villages, but made little progress overall. Contrasting this dynamism, SCs and OBCs in upper caste dominated villages and STs and OBCs in own enclaves experienced little progress between the rounds.

Were these average income changes confined to the better off or did they extend to poorer households as well? In the first round, the incidence of poverty among SCs, STs, OBCs and Muslims was lower in upper caste dominated villages than in own enclaves. Consistent with the income growth observations, the most dramatic poverty reductions appear for SCs in own enclaves and STs in upper caste dominated villages. However, in spite of modest income rises, poverty reduction among Muslims in own enclaves looks dramatic. Poorer ST households made slightly more progress than the average ST household. Consistent with the income figures, OBCs seem to have experienced limited poverty reduction between the two rounds.

Figure 2 illustrates how social identity interacted with village regime relate to one important factor endowments in rural India, namely basic education as measured by male and female illiteracy.

[Insert figure 2 about here]

For both male and female literacy, SCs and OBCs do better in their own enclaves than in UC-dominated villages and Muslims worse, in both survey rounds. No such clear pattern is discernible for STs. Among groups with low initial male literacy (SCs, STs, Muslims), we observe across the board improvements with Muslims and STs in own enclaves progressing more than those in UC dominated villages. SCs had higher and Muslims lower initial male literacy in their own enclaves. Although these observations on educational levels and progress correspond imperfectly with the income and growth patterns in Figure 1, they do provide hints of positive enclave level and growth effects for SCs. While STs in UC dominated villages experienced rapid income growth, male education does not appear to be responsible for this spur. Female STs experienced dramatic educational progress in general, while female SCs did better and female Muslims worse in own enclaves.

To sum up, in terms of the level of income and poverty in both rounds, as well as income growth and poverty reduction, UCs and STs do on average better in UC-dominated villages than anywhere else. By contrast, SCs in their own enclaves do not, on average, outperform SCs elsewhere in terms of income and poverty in round 1, but do so and apparently very strongly in round 2. Excepting poverty reduction (but not mean income growth) of Muslims in their own enclaves, the differences across village regimes for Muslims and OBCs are small. Some factor endowments – most notably SC literacy rates in SC-dominated are consistent with these patterns, but observations so far are inconclusive about how village regimes affect household welfare. We next implement the empirical strategy laid out in section II.

IV EMPIRICAL RESULTS

IVA. Estimation results

Eq. (4) is estimated for round 1 and 2 using alternative specifications where extra variables are gradually introduced, a first set primarily to ensure that key effects of interest are not locationally confounded, and a next set to investigate pathways through which enclave, proximity and oppression effects operate. The estimation results for these specifications are reported in Tables III and IV and in full in Appendix 2. The natural logarithm of real household income per capita is the dependent variable.

The first specification contains social identity dummy variables (SC, ST, MUS, OBC) with upper castes as benchmark category. We add the enclave village regime variables capturing own group land dominance (SCxDSC, STxDST, MUSxDMUS and OBCxDOBC), the enclave effect for upper castes (UCxDUC) and, finally, the social group interaction terms with upper caste dominance (SCxDUC,STxDUC, MUSxDUC and OBCxDUC). These latter interactions facilitate identification of how SCs, Muslims and OBCs perform within upper caste dominated compared to own enclaves and to the 'benchmark' other villages with the latter captured by the 'raw' social identity terms.

[Insert table III about here]

Table III is laid out to facilitate round 1 and round 2 comparisons. We discuss the enclave, proximity and oppression effects before and after introducing locational and demographic controls and proceed to address the pathways through which each of these effects operate.[29]

In columns (1) and (2) we interact the village regime variables with households' social group, first without and then with control variables added. Prior to adding agro-ecological, state and household demographic controls, it appears that residing in an upper caste dominated village not only benefits upper caste households, as suggested by the large (and significant at the 1 % level) UC enclave coefficient (UC x DUC), but also bestows sizeable benefits on OBC and SC households. The UC enclave coefficient leaps notably in size between the two rounds, but only before controls are added.

Starting with round 1 and prior to adding controls, it is evident that UC households do better than everyone else, irrespective of location, and do particularly well in UC dominated villages. OBCs do better than SCs and marginally better than STs and Muslims outside, but much better than STs and Muslims if resident in UC dominated or in OBC enclaves. In fact, OBCs are the only group that do not lose out relative to UCs in UC dominated villages. For SCs, a significant but smaller gain from residing in UC dominated villages is observed. We interpret the positive interaction terms for OBCs and SCs as the difference between positive proximity and negative oppression effects. In this first specification, the former dominate the latter.

Turning to the post-liberalisation era, we first register a general widening of identity-based disparities in favour of upper caste households. The UC enclave effect is larger and consistent with the descriptive statistics, STs do much better in UC dominated villages than anywhere else, while Muslims do better in their own enclaves (weakly significant coefficient). There is, moreover, a large and strongly positive SC enclave effect. While SCs benefitted from proximity to UCs before the reform effects started to kick in, SCs in own enclaves appear to have made significantly more progress during the post reform era.[30]

We next investigate whether the above effects are locationally confounded. UC dominated villages might be clustered in areas with greater agricultural potential and SC dominated villages in states with more progressive policies towards Scheduled Castes or in states that experienced more (or less) income growth and poverty reduction in the aftermath of the 1991 reforms; the locational disadvantage of ST dominated villages was remarked upon above.

We add three sets of controls and note that, in contrast to state dummies, the main changes occur when agro-ecological zone controls are introduced. Adding Palmer-Jones and Sen's (2003) mapping of agro-ecological zones onto Indian districts makes clear that location matters.

The results reported in column 2 show that the UC enclave coefficient sizes are sharply reduced in both rounds. Further, the interaction terms capturing OBC and SC residence in UC dominated villages turn insignificant. In contrast, the OBC enclave coefficients remain significant (shrinks in size in round 1), while the SC enclave coefficients are now significant in both rounds. The SC enclave effect remains statistically stronger and of a much larger order of magnitude in round 2. Unlike for the three broad Hindu groups, there are no discernible village regime effects for STs and Muslims.

A key insight so far is that the proximity hypothesis has merit but that the proximity gains for OBCs (but see below) and SCs (and STs in round 2) are all locationally confounded – once we control for location, the proximity and oppression effects for OBCs and SCs cancel each other out.

The main enclave coefficients are not, it turns out, locationally confounded. OBCs do better in their own enclaves in both rounds, while SCs do far better in their

own enclaves in both rounds but particularly in round 2. However, the weaker round 2 enclave effect for Muslims turns insignificant.

The precise implications of the proximity gain and of the oppression and enclave effects for income levels, growth, poverty incidence and poverty persistence are illustrated in the computations of counterfactual income, growth and poverty in subsection C below.

Our results so far suggest positive and significant enclave effects for UCs, OBCs and SCs in both rounds and no village regime effects for the other two groups. Once we control for location the OBC and SC net proximity gains that we observed to start with are wiped out. Put differently, the Hindu social groups benefit from the dominance of 'their own kind' in the village communities where they reside.[31],[32]

We next shift the analytical attention to the underlying processes at work and first study the pathways through which village regime effects operate and possible change between the rounds. We gradually control for village infrastructure, household education and household land holdings with results for the two latter reported in table IV.

[Insert table IV about here]

Following Kijima (2006), we introduce dummies for the maximum female and male education within a household where the educational categories are up to primary, middle, matriculation, higher secondary and graduate plus. A hypothesis resonating with Dercon and Krishnan's (2007) findings would be that social identity disparities – by caste, religion or tribe – should evaporate once educational

attainments are controlled for. The results reported in column 1, table IV include controls for village infrastructure and education. [33]

For both rounds, we observe a marked reduction in the raw identity coefficients (see Appendix 2) and thus in the relative disadvantage of SCs, STs, Muslims and OBCs from adding educational controls. For STs, the raw coefficient drops from -0.31 to -0.17 or by around 45 percent. For SCs, in comparison, education nets out about 33 percent of the remaining disadvantage vis-à-vis upper caste households. Our results concur with Dercon and Krishnan (2007) in suggesting that education is crucial: it is evident from the table, however, that education is only part of the solution.

Turning to the village regime effects, we observe a marginal weakening of the enclave effects for OBCs with the round 2 coefficient turning insignificant and the t-value for the round 1 coefficient slightly reduced. Overall, therefore, education sharply reduces the raw identity coefficients while leaving the village regime effects largely intact.

We next consider land holdings as potential oppression buffer or asset that may bolster enclave advantage. Starting with the raw identity terms, it is evident that controlling for household land further and substantially reduces the disadvantage of SCs and Muslims, while the effect on OBCs and STs is close to negligible. For the village regime effects, the UC enclave and the first round SC enclave effects turn insignificant once household land holdings are controlled for: it transpires that land distribution is responsible for the own enclave advantages of SCs in round 1 and the UC advantages in both rounds. For SCs in the post liberalisation era other explanations must be sought. Further, and after all controls have been added, the round 1 net proximity gain for OBCs in upper caste dominated villages resurfaces.

Notice, once more, that the raw coefficients, excepting Muslims in round 1, remain stubborn, large and statistically significant. Hence, even after location, demography, village infrastructure and key factor endowments are carefully controlled for, the raw coefficients suggest that SCs with the same resource base and attributes as others not only remain the worst off but fell further behind STs and OBCs in the post reform years. The main exception is SCs in own enclaves; the SC enclave coefficient remains large and strongly significant even after land holdings and all other controls are added and is large enough to eliminate 80 percent of the remaining disadvantage vis-à-vis UC households. Notice that Muslims also experienced a relative post-reform setback since the raw coefficient reappears as (strongly) significant in round 2.

The pathway analysis provided valuable clues about the origins of the strong enclave effects observed for UCs and SCs and less for OBCs which as noted and essentially represent the Hindu communities in our sample. The village regime effects for Muslims and STs, once location were controlled for, virtually disappeared.

Favourable land distribution holds the key to the UC and first round SC enclave advantage. What remains is to explain the persistent round 2 SC enclave effect. To obtain further clues about the underlying mechanisms, we use income share as dependent variable in four alternative specifications (e.g Benjamin et al 2011): income share from cultivation, income share from wage work, income share from business, and income share from remittances.[34] The results from these additional specifications are reported in Table V.

[**Insert table V about here**]

With a complete set of controls (including household landholdings), the cultivation share of income for SC households is higher in SC enclaves: this coefficient is large and significant at a 1% level. From the pathway analysis we already know that SC income in round 2 is much higher in such villages even after household land is controlled for. Not surprisingly, wage income share is much lower, while there is no difference in the business income share of SC households within and outside their own enclaves, nor is there a difference in the share of income from remittances. There is thus no sign that business acumen outside agriculture or higher remittances can be held responsible for the SC enclave effect. Given the strong pre-occupation with enterprise and non-farm development within the development literature, this is a surprising finding. However, and in tune with Anderson's (2011) results, the explanation needs to be sought within agriculture itself. Contrast this with ST, OBC and Muslim enclaves: For STs, there is no enclave effect on income. The cultivation share of income is higher and the wage income share lower also after household land holdings are controlled for: the business income share is also higher in ST enclaves. For Muslims and OBCs we observe similar patterns for cultivation income, while business income is significantly lower in own enclaves.

IV.B Robustness tests and auxiliary regressions

As discussed in section II, we conduct two robustness tests on our main results by replacing the dummy variables for upper caste and own group land dominance firstly with the share of village land owned by the dominant group and secondly with the fragmentation adjusted dominance measure defined by equation (3). Table VI reports the sign and the level of significance on the village regime parameters in the

specification with 'pure' control variables only (AEZs, state dummy variables and household demographic controls).

The round 1 results for these alternative specifications feature in the top half and the round 2 results in the bottom half of Table VI. 17 out of the 18 coefficients (9 per round) on the village regime variables when using the land dominance dummy are robust in terms of retaining sign and statistical significance (or insignificance, as the case may be) regardless of the dominance measure used.[35] Although it is noteworthy and reassuring that significance of coefficients is generally stronger for the more refined measures, the key results presented in Section IV.A are thus not sensitive to how dominance is measured. [36]

[Insert table VI about here]

We also implemented specifications using growth in factor endowments (land, female and male education) to explore whether upper caste dominance or own enclaves have separate effects on land or human capital accumulation in rural India. The results, reported in table A2.4, show that that while the land holdings of OBCs and Muslims in UC villages increased, these were not associated with income gains (cfr coefficients in table III, column 2). A similar observation holds for OBCs in own enclaves, but there is no change in the income coefficients between the rounds there either (table III, column 2). Interestingly and for human capital accumulation, upper caste dominance appears to inhibit the progress of other social groups. The coefficients are negative for all groups (one exception), for both males and females, with the only statistically significant coefficient observed for male OBCs. We also observe adverse enclave effects on educational progress among the non-Hindu social groups: consistent with the descriptives, these are strongly negative for females in

Muslim dominated villages: a similar and strongly negative own enclave effect is observed for male STs.

IV.C Magnitude of proximity, enclave and oppression effects

We next explore the order of magnitude of the proximity, enclave and oppression effects in terms of income, income growth, and the incidence and persistence of poverty. As noted, the proximity effect could reflect a superior quality of schools, health care and sanitation in UC dominated villages; alternatively, lower castes may emulate upper castes' stronger educational aspirations and farming practices; rich neighbours can make it less risky to adopt high yielding seed varieties since followers can absorb the good and bad experiences of wealthy early adopters (e.g. Foster and Rosenzweig 1995).

Such proximity gains could exist alongside oppression effects manifested in limitations in the access to resources or markets, a hostile school environment, exclusion from membership in the local dairy cooperative or restrictions in the access to credit schemes that facilitate response to new post reform opportunities.

To proceed, we compute counterfactual income as if the coefficients on the social identity interacted with village regime variables were equal to zero and use the coefficients from the model with pure controls reported in column 2 in Table III. In that model, the coefficient on the marginalised group dummy interacted with the UC dominated village dummy is the net effect of proximity and oppression (as explained in Section II.C). In order to disentangle the two in the simulations presented here, we set the proximity effect equal to the coefficient on UC x DUC. In other words, we perform a calculation that assumes that the estimated net effect for marginalised groups in such villages can be decomposed into a proximity effect and a remaining

oppression effect by equating the former with the enclave effect for upper-caste Hindus (see Appendix 3 for details).

[Insert table VII about here]

For round 1 and 2 income per capita and poverty, and annual income growth between the two rounds, Table VII reports, by marginalised group, actual and counterfactual figures, separately for upper-caste dominated villages and for own-group dominated villages. For the latter, counterfactual figures are based on what these variables would have been without the estimated enclave effect. For upper-caste dominated villages, three sets of counterfactual figures are reported. First, income, growth and poverty are computed as if there is no general village regime, or proximity effect (the coefficient on UC x DUC); next as if there is no group specific oppression effect (e.g. the coefficient on SC x DUC); and finally as if there is neither effect. So, for example, mean income per capita in round 1 for SCs living in UC-dominated villages is equal to 6,395 Rupees per year. Had they not benefited from the proximity effect, it would have been 5,758 Rupees; had they not suffered from oppression, it would have been 6,918 Rupees; and if neither effect were at work, it would have been 6,228 Rupees. The last figure is lower than their actual mean income, which shows that, in this case, the positive proximity effect is larger (in absolute terms) than the negative oppression effect.

The proximity effect on income of marginalised groups living in UC-dominated villages is always about 10 percent, both in round 1 and in round 2: mean income would thus have been some 10 percent lower had it not been for this effect. Since the effect on income is approximately the same size in both rounds, the growth impact is

negligible. The effect on the headcount percentage of poverty, on the other hand, depends on the group specific distribution of income in the vicinity of the poverty line. Muslims in round 1 benefited most and OBCs in round 2 least: poverty would have been 7.5 percentage points higher for the former and 4.8 percentage points higher for the latter, were it not for the proximity effect.

The group specific oppression effect on income of living in UC-dominated villages tends to be of the same order of magnitude and thus offset the proximity effect, which reflects that the net effect is usually not statistically significant, with one exception (Muslims in round 1). Income in such villages would have been 14.2 percent higher for SCs in round 2, 12.7 percent higher for OBCs in round 2, and 15.3 percent higher for Muslims in round 1. The effect on growth is pronounced, too. SCs would have experienced 1.83 instead of 1.32 percent annual growth (22.1 percent over the entire period instead of 15.6 percent) and OBCs 1.36 instead of 0.89 percent (16.0 instead of 10.7 percent), were it not for oppression. When either the oppression effect or the proximity effect dominates for income, the same effect does not always dominate in the case for poverty, which must be related to peculiarities of the PDF of income. It is worth noting, though, that poverty reduction would have been very similar in the absence of oppression – marginalised groups would have experienced about the same amount of poverty reduction as they experienced actually, because the level effect in both rounds was of the same order of magnitude.

Enclave effects in the specification used are significant only for SCs and OBCs, in both rounds. For OBCs they are of the same order of magnitude (but positive) as the oppression effects remarked on above for this group. For SCs they are much larger. Income per capita would have been 13.5 percent lower in round 1, and 25.9 percent lower in round 2, annual growth 1.47 percentage points lower (23.8 percent

less growth over the period), and poverty 7.6 and 18.0 percentage points higher in round 1 and round 2, were it not for the enclave effect. Poverty would thus have been far more persistent for SCs in own-dominated villages in the absence of this effect.

In summary, we find sizeable proximity gains to those residing in UC-dominated villages for income and poverty (but not for growth and poverty reduction), and an offsetting oppression effect of roughly the same order of magnitude. Growth for SCs and OBCs is substantially negatively affected by oppression. Enclave effects are large and positive for OBCs and especially SCs in terms of income and the absence of poverty, and for SCs in terms of growth, too.

V CONCLUDING REMARKS

Using a unique household panel data set for rural India covering the years 1993/94 and 2004/05, we have tested whether households from Scheduled Castes, Scheduled Tribes, Muslims and Other Backward Classes fare better or worse in terms of income levels when residing in villages dominated by upper castes and in villages dominated by their own group. We began by noting that the gap between Upper Caste and all other social groups widened substantially between the two panel rounds.

Our initial specification suggested a positive net gain from proximity to upper castes (e.g. Sethi and Somanathan 2010) for SCs and OBCs in round 1 and SCs, STs and OBCs in round 2 and thus that the proximity effect dominates the oppression effect. However, once we control for the agroecologically more favourable location of such villages, this net gain disappears and the proximity and oppression effects cancel each other out. A round 1 net proximity gain for OBCs resurfaces once all controls have been included, thus adding clout to the proximity hypothesis.

In order to isolate the oppression effect, we compute counterfactual household

income as if all social groups benefit equally from the advantages to the village as a whole that UC dominance brings, and find that it can be large. For instance, the income levels of SCs living in upper caste dominated villages would have been 14.2 percent higher in round 2 were it not for oppression effects, while annual income growth would have been 0.5 percentage points higher, 1.83 instead of 1.32 percent.

Put differently, while both the proximity and the oppression hypothesis have merit, neither works satisfactorily on its own. They work, moreover, in the expected opposite directions: Ignoring either through a focus on the proximity hypothesis or the oppression hypothesis alone would deprive social scientists interested in the origins of caste-based disparities in rural India of vital insights.

When focusing on income we find strong support for the *positive* enclave hypothesis for UCs, SCs and OBCs in both rounds; UCs perform much better in own dominated villages than anywhere else. SCs and OBCs also perform better in their own villages than in villages dominated by upper castes and in benchmark 'other villages'. Once location is controlled for, these village enclave effects are limited to the Hindu social groups: there are no parallel effects for STs and Muslims. In terms of income the Hindu social groups thus benefit from the dominance of 'their own kind' in the village communities where they reside.

For human capital accumulation, our findings suggest inhibiting effects of upper caste dominance on males and females from other social groups (negative signs, but only one significant coefficient) while own enclaves negatively affect educational progress for Muslim women and ST men.

We shed new light on the pathways through which welfare disparities between different social groups within and outside villages dominated by upper castes may be narrowed. Educational attainment matters, but mainly outside UC dominated villages

and outside own enclaves. The strong enclave effects for UCs in both rounds and SCs in round 1 disappear once land holdings are controlled for. The remaining gaps in the raw identify coefficients are also very substantially reduced thus underscoring that land distribution remains a key determinant of identity-based disparities in rural India. This is in contrast to Dercon and Krishnan's (2007) findings based on the ICRISAT-panel which indicated that caste-based rural disparities essentially have educational roots.

Consistent with Anderson's (2011) findings for Yadavs in Bihar and Uttar Pradesh, but in our case extending to marginalised groups below the pollution barrier, Scheduled Caste households in own dominated villages realised higher incomes in both rounds and experienced far more rapid poverty reduction between the two rounds. Our analysis of income shares suggests that the explanation for this advantage, perhaps surprisingly, is unrelated to non-farm employment or business enterprise development and is instead anchored in advantages in agricultural production: a higher return on own-account cultivation when SCs are not likely to be discriminated against in irrigation (e.g. Anderson 2011) and other markets for agricultural inputs and outputs.

Our results, based on fundamentals, provide a timely empirical corrective to accounts of sustained SC progress relative to other groups and provide an important reminder to those who, inspired by India's 'silent revolution', place great hope in the transformative potential of the democratic process whether on its own or aided by political reservations. A similar caveat applies to strong beliefs in the transformative potential of economic liberalisation. The grip of caste in rural India appears to be firmly rooted in patterns of land ownership. The exception is SCs in own enclaves

who are favourably placed for escaping this grip; the SC enclave effect remains large and strongly significant even after land holdings and all other controls are added.

REFERENCES

Akerlof, George A. and Rachel Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics* 115 (3): 715-53.

Anderson, Siwan. 2011. "Caste as an Impediment to Trade."*American Economic Journal: Applied Economics* 3(1): 239-63.

Arrow, Kenneth J. 1972. "Models of Job Discrimination." In *Racial Discrimination in Economic Life*, ed. Anthony H. Pascal. Heath, Lexington.

Banerjee, Abhijit and Rohini Somanathan. 2007. "The Political Economy of Public Goods: Some Evidence from India." *Journal of Development Economics* 87:287-314.

Banerjee, Biswajit and John B. Knight.1985. "Caste discrimination in the Indian labour market." *Journal of Development Economics* 17, 277-307.

Bardhan, Pranab.1970. "India."In *Redistribution with Growth*, ed. Hollis Chenery et al. Oxford: *Oxford University Press.*

Baulch, Bob and John Hoddinott. 2000. "Economic Mobility and Poverty Dynamics in Developing Countries." *Journal of Development Studies* 36(6), 1-24.

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

Becker, Gary S.1971. *The Economics of Discrimination* (2nd Ed.). Chicago: University of Chicago Press.

Benjamin, Dwayne, Lauren Brandt and John Giles. 2011."Did Higher Inequality Impede Growth in Rural China?" *The Economic Journal* 121(557):1281-1309.

Besley, Timothy and Robin Burgess. 2000. "Land Reform, Poverty Reduction and Growth: Evidence from India." *Quarterly Journal of Economics*, 115 (2): 389-430.

Besley, Timothy, Rohini Pande and Vijayendra Rao. 2005. "Political Selection and the Quality of Government: Evidence from South India." STICERD Research Paper DEDPS44.

Cameron, A.Colin, and Pravin K. Trivedi.2005. *Microeconometrics, Methods and Applications*. New York: Cambridge University Press.

Dercon, Stefan and Pramila Krishnan. 2007. "Caste revisited: Consumption Dynamics and Exclusion in South-Indian villages 1975-2005." Mimeo. Oxford: University of Oxford.

Desai, Sonalde, Amaresh Dubey, B.L. Joshi, Mitali Sen, Abusaleh Shariff and Reeve Vanneman. 2009. "India Human Development Survey: Design and Data Quality." Technical Paper No. 1. NCAER and University of Maryland.

Deshpande, Ashwini. 2001. "Caste at Birth? Redefining Disparity in India."*Review of Development Economics* 5(1): 130-44.

Deshpande, Ashwini. 2011. *The Grammar of Caste: Economic Discrimination in Contemporary India*. New Delhi: Oxford University Press.

Dreze, Jean, Peter Lanjouw and Naresh Sharma. 1999. "Economic Development in Palanpur, 1957-93." *In Economic Development in Palanpur over Five Decades*, eds Peter Lanjouw and Nicholas Stern. Oxford: Clarendon Press.

Do, Quy-Toan and Lakshmi Iyer (2010): Geography, Poverty and Conflict in Nepal, *Journal of Peace Research*, 47(6).

Dumont, Louis.1970. *Homo hierarchicus. The caste system and its implications*.London: Paladin Press.

Foster, Andrew and Mark R. Rosenzweig. 1995. "Learning by doing and learning from others: human capital and technical change in agriculture." *Journal of Political Economy* 103: 1176-1209.

Fuller, Chris J. 1996. *Caste Today*. Oxford: Oxford University Press.

Gang, Ira N., Kunal Sen and Myeung-Su Yun.2008a. "Poverty in Rural India: Caste and Tribe." *Review of Income and Wealth* 54 (1): 50-70.

Gang, Ira N., Kunal Sen and Myeung-Su Yun.2008b. "Was the Mandal Commission Right? Living Standard Differences between Backward Classes and Other Social Groups in India." Discussion Paper 3453. Bonn: Institute for the Study of Labor (IZA).

Guha, Ramachandra. 2007. *India after Gandhi: The History of the World's Largest Democracy*. New Delhi: Pan MacMillan.

Hoff, Karla, Mayuresh Kshetramade and Ernst Fehr. 2009. "Caste and Punishment: The Legacy of Caste Culture in Norm Enforcement." World Bank Policy Research Working Paper No. 5040.

Iversen, Vegard and Raghavendra, P.S. 2006. "What the signboard hides: Food, caste and employability in small, South-Indian eating places." *Contributions to Indian Sociology* 41(3):311-41.

Iversen, Vegard. 2012. "Caste and Upward Mobility", in *The Oxford Handbook of the Indian Economy*, ed. Chetan Ghate. New York: Oxford University Press.

Jaffrelot, Christophe. 2003. *India's Silent Revolution – The Rise of the Low Castes in North Indian Politics*. Ranikhet: Permanent Black.

Kapur, Devesh, Chandra Bhan Prasad, Lant Pritchett and D. Shyam Babu. 2010. ”Rethinking Inequality: Dalits in Uttar Pradesh in the Market Reform Era.” *Economic and Political Weekly*, 46(35): 39-49.

Kijima, Yoko. 2006. “Caste and Tribe Inequality: Evidence from India, 1983-1999.” *Economic Development and Cultural Change* 54(2): 369-404.

Kingdon, Geeta. 1998. “Does the Labour Market Explain Lower Female Schooling in India?” *Journal of Development Studies* 35(1): 39-65.

Madehswaran, S. and Paul Attewell . 2007. “Caste Discrimination in the Indian Urban Labour Market: Evidence from the National Sample Survey.” *Economic and Political Weekly*, 13 October, 4146-4153.

Madsen, Stig Toft. 1991. “Udupi hotels: A case of ethnicity based Non-Virulent Development.” Presented at First Nordic Conference on South-Asian studies.

Menon, Dilip M. 1994. *Caste, nationalism and communism in South-India – Malabar 1900-1948*. Cambridge: Cambridge University Press.

Nambissan, Geetha B. 1996. “Equity in Education? The Schooling of Dalit Children in India.” Discussion Paper no. 15. New Delhi: United Nations Development Programme.

Palmer-Jones, Richard and Kunal Sen. 2003. “What Has Luck Got To Do With It? A Regional Analysis of Poverty and Agricultural Growth in Rural India.” *Journal of Development Studies* 40:1-31.

Panagariya, Arvind. 2008. *India: The Emerging Giant*. Oxford: Oxford University Press.

Pande, Rohini .2003. “Can Mandated Political Representation Increase Policy Influence for Disadvantaged Minorities? Theory and Evidence from India.” *American Economic Review,* 93 (4): 1132-1151.

Peyton Young, Hobart(2001): "The dynamics of conformity." In Social Dynamics, eds Steven N. Durlauf and H. Peyton Young. Cambridge (Mass): *The MIT Press*.

Pinstrup-Andersen, Per and Satoru Shimokawa. 2006. "Rural Infrastructure and Agricultural Development." Presented at World Bank Conference on Development Economics, Tokyo, Japan, May 29-30, 2006.

Rosenzweig, Mark R. 2003. "Payoffs from Panels in Low-Income Countries: Economic Development and Economic Mobility." *American Economic Review* 93(2), 112-117.

Sen, Kunal. 2007. "Why did the elephant start to trot? India's growth acceleration re-Examined." *Economic and Political Weekly*, October 27-November 2, 37-49.

Sethi, Rajiv and Rohini Somanathan. 2010. "Group Identity and Social Mobility in India." Paper presented at the Conference on Social Policy in India, University of Warwick, 2-3rd July 2010.

Shah, Ghanshyam Harsh Mander, Sukhdeo Thorat, Satish Deshpande and Amita Baviskar. 2006. *Untouchability in rural India.* New Delhi: Sage Publications.

Singh, Harjinder. 1984. *Caste among non-Hindus in India*. New Delhi: National Publishing House.

Srinivas, M. N. (1955): The Social System of a Mysore Village, in M. McKim (ed.): Village India, *University of Chicago Press*.

Thorat, Sukhdeo and Paul Attewell. 2007. "The Legacy of Social Exclusion: A Correspondence Study of Job Discrimination in India." *Economic and Political Weekly*, 13, October.

Table I: Number of land and population dominated villages by social group

	Largest land-holding group in village						
	SC	ST	OBC	MUS	UC	OTH	Total
Largest population group in village:							
Scheduled Castes (SC)	**24**	2	12	1	25	4	68
Scheduled Tribes (ST)	0	**65**	3	0	2	0	70
Other Backward Classes (OBC)	1	0	**196**	3	25	10	235
Muslims (MUS)	0	0	2	**35**	4	2	43
Upper Castes (UC)	0	1	1	0	**223**	6	231
Others and none (OTH)	2	0	8	0	18	**1**	9
Total	27	68	222	39	297	26	**679**

Source: HDPI panel, authors' calculations

Notes: Figures are number of panel villages in which the row social groups are the largest population group and the column social groups own the largest land share. The category "others" consists of villages in which either an unclassified group or no single group is land- or population-dominant.

Table II: Number of households by social group and village regime

	Land dominant social group						
	SC	ST	OBC	MUS	UC	OTH	Total
Social group of households:							
Scheduled Castes (SC)	**222**	68	694	109	1,040	119	2,252
Scheduled Tribes (ST)	23	**552**	141	21	95	3	835
Other Backward Classes (OBC)	86	169	**1,608**	64	852	130	2,909
Muslims (MUS)	52	10	130	**337**	145	25	699
Upper Castes (UC)	44	61	381	29	**1,810**	91	2,416
Total	427	860	2,954	560	3,942	368	9,111

Source: HDPI panel, authors' calculations.

Table III Estimation results of the effects on income of social identity, village regime and locational and demographic controls

Model:	Village regime terms (1)		Plus controls (2)	
	Round 1	Round 2	Round 1	Round 2
Social identity:				
HH is SC	-0.383***	-0.506***	-0.380***	-0.450***
	(-8.21)	(-10.15)	(-8.58)	(-9.45)
HH is ST	-0.316***	-0.461***	-0.315***	-0.372***
	(-4.37)	(-5.67)	(-4.89)	(-5.35)
HH is OBC	-0.296***	-0.310***	-0.230***	-0.241***
	(-4.85)	(-4.40)	(-4.10)	(-3.92)
HH is MUS	-0.294***	-0.445***	-0.207***	-0.323***
	(-5.34)	(-5.96)	(-3.68)	(-4.71)
Village regime variables:				
SC x DSC	0.037	0.264***	0.145**	0.300***
	(0.49)	(2.74)	(2.04)	(3.06)
ST x DST	-0.088	-0.050	-0.018	0.003
	(-1.16)	(-0.59)	(-0.27)	(0.05)
OBC x DOBC	0.167***	0.113*	0.105**	0.094*
	(2.88)	(1.74)	(1.97)	(1.75)
MUS x DMUS	-0.023	0.164*	-0.016	0.124
	(-0.29)	(1.74)	(-0.20)	(1.37)
UC x DUC	0.198***	0.288***	0.105**	0.109**
	(3.84)	(5.14)	(2.22)	(2.13)
SC x DUC	0.088**	0.139***	0.0264	-0.024
	(2.04)	(3.01)	(0.60)	(-0.54)
ST x DUC	-0.020	0.260**	-0.023	0.103
	(-0.15)	(2.10)	(-0.20)	(0.341)
MUS x DUC	0.013	0.142	-0.037	0.026
	(0.16)	(1.46)	(-0.46)	(0.29)
OBC x DUC	0.176***	0.156**	0.032	-0.0105
	(2.85)	(2.32)	(0.55)	(-0.18)
Controls:				
Household composition	No	No	Yes	Yes
Agro-ecological zones	No	No	Yes	Yes
State dummy variables	No	No	Yes	Yes
R squared (overall)	0.0667	0.1065	0.2124	0.2837
N	9108	9108	9108	9108

Source: HDPI-I ("round 1") and II ("round 2") surveys, panel households only; authors' calculations. Notes: Dependent variable is the natural logarithm of annual per capita household income in constant 1993/94 prices, with round 2 figures converted using NSSO state-specific rural CPIs. Random effects, with standard errors that are robust to heteroskedasticity and clustering within villages. ***, ** and * denote significance at 1, 5 and 10 percent respectively; robust t-statistics are in parentheses. Demographic controls are the sex of the household head, number of boys aged 0-5, girls 0-5, boys 6-14, males 15-19, females 15-19, males 20-24, females 20-24, males 25-49, females 25-49, males 50-59, females 50-59, males 60 and older, and females 60 and older. See table A2.1 in Appendix 2 for the full specification.

Table IV Estimation results of the effects on income of social identity, village regime and additional controls: village infrastructure, household education and land

Controls added:	Plus education (hh)		Plus land (hh)	
	Round 1	Round 2	Round 1	Round 2
Social identity:				
HH is SC	-0.255***	-0.316***	-0.166***	-0.248***
	(-6.17)	(-7.18)	(-4.27)	(-6.06)
HH is ST	-0.177***	-0.232***	-0.157***	-0.202***
	(-2.85)	(-3.49)	(-2.73)	(-3.08)
HH is OBC	-0.154***	-0.158***	-0.136***	-0.128***
	(-2.85)	(-2.83)	(-2.81)	(-2.62)
HH is MUS	-0.119**	-0.216***	-0.070	-0.138**
	(-2.17)	(-3.31)	(-1.38)	(-2.22)
Village regime:				
SC x DSC	0.129**	0.249***	0.067	0.202**
	(2.01)	(2.70)	(1.09)	(2.41)
ST x DST	-0.007	0.009	0.012	-0.008
	(-0.12)	(0.14)	(0.21)	(-0.14)
OBC x DOBC	0.100*	0.074	0.067	0.030
	(1.94)	(1.50)	(1.51)	(0.68)
MUS x DMUS	0.005	0.15*	-0.047	0.087
	(0.07)	(1.80)	(-0.74)	(1.15)
UC x DUC	0.120***	0.108**	0.032	0.026
	(2.66)	(2.16)	(0.79)	(0.56)
SC x DUC	0.015	-0.04	0.044	-0.018
	(0.34)	(-0.98)	(1.18)	(-0.46)
ST x DUC	-0.006	0.124	0.003	0.138
	(-0.05)	(1.24)	(0.04)	(1.51)
MUS x DUC	-0.004	0.050	0.062	0.026
	(-0.05)	(0.54)	(0.83)	(0.30)
OBC x DUC	0.039	-0.009	0.11**	0.023
	(0.69)	(-0.18)	(2.25)	(0.49)
Controls:				
Household composition	Yes	Yes	Yes	Yes
Agro-ecological zones	Yes	Yes	Yes	Yes
State dummies	Yes	Yes	Yes	Yes
Village infrastructure	Yes	Yes	Yes	Yes
Household education	Yes	Yes	Yes	Yes
Household land	No	No	Yes	Yes
R squared (overall)	0.2702	0.3413	0.4255	0.4182
N	9108	9108	9108	9108

Source and Notes: as for Table III.

Additional notes: Education variables are dummy variables used as controls for the highest level of male and female education in the household. Land refers to controls for the logarithm of owned household land measured in acres, and the logarithm of irrigated household land measured in acres. Village size is captured by village population (logarithm) and total village land. The village infrastructure controls are the presence within the village of a busstop (1), or within its vicinity of a railway station (2), medical clinic (3), schools, and if so, at which level of education (4), or a market/mandi (5), as well as the type of road (footpath only, kutcha road, pucca road) that leads to the village (6). The full specification is reported in table A2.2 in Appendix 2.

Table V: Round 2 estimation results of the effects on income share of social identity, village regime with full set of controls

	Agriculture	Wage	Business	Remittances
Social identity:				
HH is SC	-0.137***	0.196***	-0.030**	-0.014*
	(-7.15)	(9.60)	(-2.30)	(-1.71)
HH is ST	-0.098***	0.186***	-0.061***	-0.013
	(-3.64)	(6.13)	(-3.44)	(-1.33)
HH is OBC	-0.014	0.008	0.042**	-0.013
	(-0.60)	(0.32)	(2.17)	(-1.38)
HH is MUS	-0.098**	-0.009	0.111***	0.005
	(-3.29)	(-0.27)	(3.32)	(0.33)
Village regime:				
SC x DSC	0.116***	-0.151***	0.023	0.006
	(3.84)	(-3.16)	(1.25)	(0.41)
ST x DST	0.081***	-0.112	0.040**	-0.001
	(2.57)	(-3.39)	(2.55)	(-0.14)
OBC x DOBC	0.075***	-0.034	-0.049***	0.006
	(3.24)	(-1.34)	(-2.76)	(0.81)
MUS x DMUS	0.066	0.0107	-0.070*	-0.009
	(1.58)	(0.24)	(-190)	(-0.52)
UC x DUC	0.080***	-0.060***	-0.013	0.005
	(3.63)	(2.77)	(-1.04)	(0.67)
SC x DUC	-0.008	-0.066	0.002	0.006
	(-0.50)	(-0.33)	(0.26)	(0.98)
ST x DUC	0.036	-0.050	0.018	-0.005
	(-0.05)	(-1.15)	(0.325)	(-0.37)
MUS x DUC	-0.106	0.071	-0.106	-0.027
	(-0.25)	(1.30)	(-0.24)	(-1.52)
OBC x DUC	-0.024	0.016	-0.009	0.012
	(0.69)	(0.56)	(-0.47)	(1.43)
Controls:				
Household composition	Yes	Yes	Yes	Yes
Agro-ecological zones	Yes	Yes	Yes	Yes
State dummies	Yes	Yes	Yes	Yes
Village infrastructure	Yes	Yes	Yes	Yes
Household education	Yes	Yes	Yes	Yes
Household land	Yes	Yes	Yes	Yes
R squared (overall)	0.3375	0.2679	0.0842	0.1278
N	9108	9108	9108	9108

Source and Notes: as for Table III.

Additional notes: Education variables are dummy variables used as controls for the highest level of male and female education in the household. Land refers to controls for the logarithm of owned household land measured in acres, and the logarithm of irrigated household land measured in acres. Village size is captured by village population (logarithm) and total village land. The village infrastructure controls are the presence within the village of a busstop (1), or within its vicinity of a railway station (2), medical clinic (3), schools, and if so, at which level of education (4), or a market/mandi (5), as well as the type of road (footpath only, kutcha road, pucca road) that leads to the village (6). The full specification is reported in table A2.2 in Appendix 2.

Table VI: Qualitative summary of robustness tests

Main dominance measure:	*Main dominance measure results*	*Land percentage of largest land holding group in village*	*Dominance-adjusted Herfindahl index (eq. 3)*
Round 1			
SC x DSC	++	++	++
ST x DST	Ns	Ns	Ns
OBC x DOBC	++	Ns	Ns
MUS x DMUS	Ns	Ns	Ns
UC x DUC	++	+++	+++
SC x DUC	Ns	Ns	Ns
ST x DUC	Ns	Ns	Ns
MUS x DUC	Ns	Ns	Ns
OBC x DUC	Ns	Ns	Ns
Round 2			
SC x DSC	+++	+++	+++
ST x DST	Ns	Ns	Ns
OBC x DOBC	+	+++	+++
MUS x DMUS	Ns	Ns	Ns
UC x DUC	++	+++	+++
SC x DUC	Ns	Ns	Ns
ST x DUC	Ns	Ns	Ns
MUS x DUC	Ns	Ns	Ns
OBC x DUC	Ns	Ns	Ns

Notes: +++, ++, + indicates positive coefficient significant at 1, 5 and 10% respectively, ---, --, - indicates negative coefficient significant at 1, 5 and 10% respectively, Ns indicates not significant, all in the specification with social group, village regime, agro-ecological zones, state dummies and household demographic composition variables.

Table VII: Actual and counterfactual annual per capita income (in 1993/94 Rupees), growth (% per year) and poverty (%) without village regime effects by social group

	Scheduled Castes		***Scheduled Tribes***		***Other Backward Classes***		***Muslims***	
	Round 1	Round 2	Round 1	Round 2	Round 1	Round 2	Round 1	Round 2
Upper-caste dominated villages								
Mean income per capita								
Actual	6,395	7,391	6,760	8,905	8,309	9,200	6,626	7,915
Counterfactual – without proximity effect	5,758	6,628	6,086	7,985	7,480	8,249	5,965	7,098
Counterfactual – without oppression effect	6,918	8,443	7,683	8,959	8,938	10,367	7,637	8,600
Counterfactual – without proximity and oppression effects	*6,228*	*7,571*	*6,917*	*8,034*	*8,046*	*9,297*	6,876	*7,712*
Growth in mean income per capita (% per year between 1994 and 2005)								
Actual	-	1.32	-	2.54	-	0.93	-	1.63
Counterfactual – without proximity effect	-	1.29	-	2.50	-	0.89	-	1.59
Counterfactual – without oppression effect	-	1.83	-	1.41	-	1.36	-	1.09
Counterfactual – without proximity and oppression effects	-	*1.79*	-	*1.37*	-	*1.32*	-	1.05
Poverty headcount (%)								
Actual	43.9	35.5	46.3	33.7	30.8	27.7	35.9	33.1
Counterfactual – without proximity effect	49.5	42.2	51.6	41.1	35.6	32.4	43.4	40.0
Counterfactual – without oppression effect	38.9	29.9	38.9	31.6	27.2	22.5	29.0	26.9
Counterfactual – without proximity and oppression effects	*45.6*	*34.4*	*46.3*	*41.1*	*32.2*	*27.5*	33.8	*36.6*
Own-group dominated villages								
Mean income per capita								
Actual	5,954	9,842	5,331	5,805	8,158	9,187	6,553	8,231
Counterfactual – without enclave effect	5,151	7,291	*5,427*	*5,788*	7,345	8,362	*6,658*	*7,271*
Growth in mean income per capita (% per year between 1994 and 2005)								
Actual	-	4.68	-	0.78	-	1.09	-	2.09
Counterfactual – without enclave effect	-	3.21	-	*0.59*	-	1.19	-	*0.80*
Poverty headcount (%)								
Actual	52.3	29.7	50.9	47.1	34.8	30.2	51.0	37.1
Counterfactual – without enclave effect	59.9	47.7	*50.4*	*47.3*	40.4	36.1	*50.1*	*45.1*

Notes: counterfactual figures are all based on counterfactual income computed for each household in villages land dominated by indicated group, using coefficients from the round 1 and round 2 regressions of the natural logarithm of income on village regime and social identity variables, controlling for agro-ecological zones, state dummies, and household demographic characteristics, as reported in column 2, table III and in full in Appendix 2. Figures in italics are based on coefficients insignificant at the 10 percent level.

Figure captions:

Figure 1: Mean per capita household income (in 1993/94 Rupees) and poverty headcount (proportion) by social group, round and village regime

Source: HDPI panel, authors' calculations

Notes: Poverty is the share of the indicated sub-sample with income below the NSSO state-specific rural poverty lines.

Figure 2: Male and female illiteracy by social group, round and village regime

Source: HDPI panel, authors' calculations

Notes: Figures are averaged across all *households* in the sub-sample indicated, and are based on the highest level of educational achievement in the household, i.e. on households of which not a single (female or male, as appropriate) member is literate.

[+] We would like to thank Farzana Afridi, Siwan Anderson, Sanghamitra Das, Ashwini Deshpande, Chetan Ghate, Stefan Jonsson, Anirban Mitra, Rinku Murgai, Richard Palmer-Jones, Indira Rajaraman, Debraj Ray, Kunal Sen and two anonymous referees for helpful comments and suggestions. We are also indebted to Richard Palmer-Jones for sharing the classification of agro-ecological zones used in this paper.

[1] Apart from Anderson (2011), the caste dominance concept has been applied in economic studies by, among others, Besley, Pande and Rao (2005), Dercon and Krishnan (2007), and Do and Iyer (2010).

[2] Caste may refer to *jati* (sub-caste) or to the more general *varna*, the latter comprising four broad occupational groups with *Brahmins* at the top followed by *Kshatriyas* (warriors), *Vaishyas* (traders and merchants) and *Shudras* (manual workers and craftspersons) at the bottom. SCs may be portrayed as a subset of the Shudras or a separate category. Their main distinguishing characteristic is a particularly degrading ('polluting') traditional occupation.

[3] The criteria for Scheduled Tribe classification are (i) tribal origin; (ii) primitive ways of life and habitation in remote and less accessible areas; (iii) general backwardness in all respects (Pande 2003, 1138).

[4] Examples from the recent past include caste demarcators in how people dressed and spoke and what they were allowed to do. In 19th Century Kerala, "*when a Namboodiri Brahmin approached, a Paraiya labourer had to cry out in advance, lest the sight of him pollute his superior*" (Guha 2007, 287). Also in Kerala and during conversations with a person of higher caste, members of lowly ranked castes were expected to use debasing words to describe themselves (Menon 1994,19). Nambissan (1996) presents

historical evidence of how Scheduled Caste children, while permitted to attend school, could be denied entry to the classroom.

[6] The issue was first addressed by the Other Backward Class Commission, appointed by Prime Minister Nehru, and later and more decisively by the Mandal Commission (1978-80). The latter's recommendations, extending reservation benefits to OBCs, were declared constitutionally legitimate in 1992.

[7] This is in contrast to the widespread changes in social practices in Western and Eastern Uttar Pradesh reported by Kapur et al (2010). However and unlike Kapur et al (2010), we focus on a fundamental, namely household income.

[8] Existing studies make use of nationally representative cross-sectional data and Blinder Oaxaca or alternative decomposition techniques to quantify the disadvantage associated with Scheduled Caste, Scheduled Tribe or religious identity (e.g. Kijima 2006; Gang et al 2008a). Dercon and Krishnan (2007) use the ICRISAT household panel but their analysis is limited to 204 households from six villages and two states. Lower educational attainment accounts for the slower standard of living improvements of SC/STs.

[9] As Gang et al (2008a) note, present labour market disadvantage may not reflect labour market discrimination but that cross-section analysis picks up pre-market variation in the quality of education received. While recent studies of upper end labour markets use field experiments to tackle such hurdles to identification, inferences are limited to discrimination at the point of labour market entry (e.g. Thorat and Attewell 2007).

[10] Exceptions include Dreze and Kingdon (2001) who find that rural Scheduled Caste children have an 'intrinsic disadvantage' and a lower chance of attending school even

after household wealth, parental education and motivation and school quality are controlled for. See Hoff et al (2009) for an experimental explanation for the greater economic vulnerability of lower castes.

[11] We hence consider the liberalisation not as a discrete historical event but an ongoing process with cumulative impacts over time. Neither GDP growth, growth in the services sector nor private sector investment had picked up by the time the first panel round (1993/94) was completed. For supportive evidence and more comprehensive accounts of India's growth turnaround, see Sen (2007) and Panagariya (2008).

[12] Banerjee and Somanathan (2007) find that parliamentary constituencies with a concentration of Brahmins had better access to schools and piped water in 1971.

[13] A full description of the variables, summary statistics including comparison with other major India surveys, and an exposition of the sampling methodology can be found in Desai et al. (2009).

[14] These point estimates are close between the data sources, as is the implied reduction in rural poverty. The NCAER income data imply a rural poverty decline of 9.3 percentage points, and the NSSO expenditure data of 8.7 points.

[15] See Singh's (1984) account of caste among non-Hindus. Among Muslims, Fuller (1996) and other contributors to the same volume contend that while caste-like arrangements are common, few admit to their existence. See also Appendix 1.

[16] States included in the panel are Andhra Pradesh, Bihar (+ Jharkand in round 2), Gujarat, Haryana, Himachal Pradesh, Kerala, Madhya Pradesh (+ Chhattisgarh in round 2), Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh (+ Uttarakhand in round 2) and West Bengal. Recontact details were largely lost in Assam due to a flood and in Karnataka because of human error. The 12 remaining

households in Assam do not feature in the final analysis because of lack of information on Agro-Ecological Zones (see below).

[17] See Table 1 in Appendix in Desai et al (2009) which reports the proportions of the panel household sample in round 2 and those of the refresher sample in categories of age (8 categories), gender (2), individual education (6), social group (6), place of residence (4), maximum adult education (6), and income (6). The absolute differences between the proportions of the two samples (38 comparisons in total) range from 0.04 to 5.28 percentage points, with a mean value of 1.20 and a median of 0.56 percentage points.

[18] The p-value corresponding to the null hypothesis that income is not associated with panel inclusion is equal to 0.937.

[19] These results are not presented or further discussed here, but are available on request.

[20] See the extensive literature review on land markets in Dreze et al (1999).

[21] Details are available from the authors.

[22] Throughout income is per capita per annum and in constant 1993/94 prices, converted using NSSO state-specific rural CPIs.

[23] This is equivalent to estimating (4) with one modification: replace $\beta_5 UC$xDUC with $\beta_5 DUC$, which we did in a previous version of the paper. All coefficients are unaffected by this modification, apart from the γ–parameters, the effects on income for other social groups of living in a UC-dominated village, which are then net of the approximated proximity effect, and can thus be thought of as an oppression effect.

[24] The relevance of locational disadvantage, which corresponds highly imperfectly with state boundaries, for poverty (and inequality) in rural India is extensively documented by Palmer-Jones and Sen (2003).

[25] Echoing Bardhan's (1970) assertion that redistributive reforms have not been implemented with sincerity.

[26] The general inactivity of land markets emphasised in footnote 20 adds further impetus to Anderson's (2011) claim.

[27] Their map (Palmer-Jones and Sen (2003,14-15)) divides India into 19 agro-ecological zones where careful classifications of land surface capture initial conditions that indicate agricultural productivity potential. The zones are classified by variation in soil types, rainfall patterns, altitude, whether coastal and other factors that affect this potential. Two examples of these zone definitions are zone 7: Deccan Plateau of Telangana and Eastern Ghats, hot semi-arid eco-region with red loamy soils – GP 90-150 d. and zone 5: Central (Malwa) highlands, Gujarat plains and Kathiarwar peninsular, hot arid ecoregion with medium and deep black soils and GP 90-150 d.

[28] Other descriptive statistics for this panel including mean household income by state, land holdings, levels of education (of the household head), occupation and real household income per capita for different social groups and show a close correspondence between a priori expectations and summary statistics. Marginalised social groups own less land and are less educated than others. 41% of SC households and 48 % of Muslim households have their own land; the figures for STs, OBC and UCs are 70%, 63% and 81%, respectively. Consistent with Kijima (2006), marginalised communities also appear to receive lower returns on their human capital.

[29] We report the 'raw' social identity coefficients without any controls in table A2.1 in Appendix 2. These coefficients suggest that the disparity between upper castes and each of the other social groups widened during the reform years.

[30] The responses of the 'raw' identity and village regime coefficients to the step-wise introduction of each of the three sets of 'pure' controls may be gauged in full in table A2.1 in Appendix 2.

[31] The results in our first raw regressions appear to be inconsistent with the observation of STs making less progress than SCs in the post liberalisation era (e.g. Iversen 2012). We observe that STs do better than SCs in non-UC and non-(SC or ST) dominated villages while SCs do better than STs in UC dominated villages in round 1 and better than STs in UC dominated villages and in particular in their own SC enclaves in round 2. This, if anything, provides a more nuanced picture than offered elsewhere. We also, in response to a request from a reviewer, included villages dominated by OBCs along with social group interactions to explore whether OBC dominance affected SCs or other social groups differently. The only insight on offer is that ST households in such villages fare notably worse in round 2.

[32] Sample size limitations prevent the exploration of these enclave effects at the jati-level.

[33] As can be seen in Appendix 2, adding the village infrastructure controls detailed in the note to table IV has close to negligible effects on the raw identity and village regime coefficients. On the face of it and contrary to received wisdom (e.g Pinstrup-Andersen and Shimokawa 2006), the scope for reducing identity based disadvantage by improving village infrastructure appears more limited than expected.

[34] The full set of results is available on request.

[35] The exception is the coefficient on OBC x DOBC in round 1, which is no longer significant when alternative dominance measures are used.

[36] In response to a referee request, we implemented separate regressions by social groups with results reported in table A2.3. In spite of the much fewer observations,

the enclave results are retained for OBCs and SCs (round 2). For UCs, enclave coefficients turn insignificant in both rounds - at the outset a source of concern. Given that the latter could simply reflect the much smaller sample, we ‘compensate’ for the loss of observations by replacing the dominance dummy with the two more refined dominance measures in the UC regressions. For these two more refined measures, the significance of the round 1 UC enclave coefficient is restored (at the 5 % level), while the round 2 coefficient is borderline insignificant. We hence conclude that our UC enclave results are robust.

APPENDIX 1: Construction of variables to capture upper caste dominance.

The village and household questionnaires contain data on three classifications of social groups, firstly and most disaggregated by *jati* [and name of tribe] (C1) (for Hindus, Muslims, Sikhs and STs), secondly by five broad categories (C2), namely *Brahmin*, OBC (Other backward classes), SC (Scheduled Caste), ST (Scheduled Tribe) and Other and finally by eight religious categories (C3), Hindu, Muslim, Christian, Sikh, Buddhist, Jain, Tribal and Other. The village questionnaire also contains information on the most (upto eight) numerous *jatis*, the percentage of the village population each of these *jatis* represent, and the percent of village land owned by each of these same *jatis*.

The oppression hypothesis is founded on the notion of (upper) caste dominance. If restricted to ritual rank, a simple and narrow definition would be to limit the upper caste label to *Brahmins*. Notions of upper caste advantage (and dominance) do, however, stretch beyond this top layer of the varna hierarchy.[1,2] A pragmatic alternative would be to add the "Other" category from the household questionnaire; the combination *Brahmin* (C2) plus "Other (C2)" and Hindu (C3) would then represent a broad definition of upper or forward caste *Hindus*.

This latter option has important limitations; Firstly, the exclusive focus on Hindus would miss out on social groups who may be in a position to wield considerable power and influence but who belong to a different faith. To illustrate, some of the numerically important *jatis* in our panel transcend religious boundaries; in Punjab there are significant numbers of Sikh and Hindu *Jat* households and Sikh and Hindu Dalit households with inter-caste violence involving *Jat* and Dalit Sikhs.[3] For Muslims and noted in footnote 15 in the main text, Fuller (1996) and others in the same volume contend that while caste-like arrangements are common, few within the Muslim community admit to their existence.[4] In spite of social ranks among Muslims, the less accurate reporting of the social groups that Muslim panel households belong

[1] In addition, the prevalence of *Brahmin* households varies across regions.

[2] Even among *Brahmins* there are, of course, more fine-tuned internal rankings – *Gouda Saraswath* or *Konkani* Brahmins, who are fish eating residents of Karnataka's Coastal belt, have lower social status locally than the strictly vegetarian *Madhwa* or *Udupi Brahmins*.

[3] See http://hinduonnet.com/fline/fl2013/stories/20030704002703900.htm. Punjab is also the state with the highest percentage of Scheduled Castes in its population (28.9 % according to Census of India 2001). See Jodhka's (2004) discussion on Sikhism and caste.

[4] Jeffrey et al (2007: 43) note how 'during the pre-colonial era there were marked divisions between a very small, upper caste Muslim elite and other Muslims castes, such as weavers, carpenters and barbers'.

to, left us with no other option but to define Muslim households by their religion alone. A similar strategy was adopted for Scheduled Tribes. Although the tribe a household belongs to is accurately reported, ethnographic evidence is not supportive of local hierarchies; STs thus features as a single social category in our analysis.

Secondly, the process of "de-Sanskritisation", whereby social groups lobby to downgrade their official status in order to avail of reservation benefits implies that the definitions of forward castes that anthropologists and sociologists, informed by careful field observations, subscribe to, are increasingly out of tune with official and survey data social group categories. The implementation of the Mandal Commission's (1978-80) recommendations added fresh impetus to reservations as political battleground and in the present political climate, it is not unusual to interpret the absence of 'backward' status as evidence of a social group's lack of political clout. Important groups that have acquired OBC status, include the 'clean-caste' *Vokkaligas*, the dominant peasant caste in Central and Southern Karnataka (e.g. Srinivas 1978; Epstein et al. 1998), the ritually superior *Lingayats* in the same state (Bayly 1999; 294) and more recently the *Jats* in Uttar Pradesh (e. g. Jeffrey 2001) and Rajasthan; official status is therefore, in key instances and increasingly, a reflection of political opportunism aimed at placating important vote banks with the unfortunate side effect of weakening the reliability of official status as indicator of ritual status.[5]

Other variations in caste status are found at the lower end: *Nuniyas* and *Dhanuks*, who are OBCs in Uttar Pradesh, have Scheduled Caste status in West-Bengal. *Dhobis* (washermen), have SC status in some states but not in others. For *jatis* traditionally concentrated in the most degrading occupations, like leatherworkers (e.g. *Chamars*) and sweepers (e.g. *Balmikis*), SC status is less variant to state boundaries.

Further, social groups that are not OBC, SC or ST should not necessarily be treated as upper or forward castes for analytical purposes. There are intermediate social groups in many regions for whom a more fine tuned distinction is desirable. *Rods*, an important agricultural caste in Haryana, is classified as 'other' and thus forward officially as well as in the household questionnaire; this does not square with anthropological field observations (Prem Chowdhry, pers comm.). Further, and in tune with the Mandal commission's view and report, important agricultural castes

[5] While de-Sanskritisation so far has tended to involve attempts to gain OBC-status, recent agitations by the *Gujjar*-community based on comparisons with the *Meena* community in Rajasthan aimed to downgrade their official status from OBC to ST. Similarly, in an article on UP politics, the Deccan Herald (4 March 2008), listed a number of groups whose official status were proposed 'downgraded' from OBC to SC.

such as the *Kurmis* of North and *Kunbis* of Central India do not enjoy the same local stature as *Jats* and *Marathas*, respectively (Singh 1992; 41 and Report of the Backward Class Commission, p.56 as cited in Jaffrelot 2003; 323). For the former two, the OBC classification is therefore appropriate.

In our interpretation of upper caste which is informed by anthropological observations, we adhere to ritual rank as far as the top and bottom layer is concerned, but disconnect, whenever appropriate and for reasons mentioned above, from *official* categories for the more fluid middle layer. While this imposes an additional work burden, it is important to distinguish our small-scale endeavour from past efforts to develop comprehensive caste rankings for rural India. British colonial administrators have subsequently been caricatured for believing in the possibility of such a task which at the time paved the way for an obsession with caste and jati among late Victorian data collectors (Bayly 1999, chapter 3). For North-India, our classification of the most important and by far the most numerous groups (and households in our panel) is consistent with the Mandal Commission's views and according to which the following broad groups should be treated as forward or upper castes; *Brahmins* (including Bhumihars) *Rajputs*, *Kayasthas*, *Jats*, *Marathas*, *Vaishyas*/*Banias* (Jaffrelot 2003; 323).

An informed reader will notice the inclusion of cultivating castes like North-Indian *Jats* along with the conspicuous absence of similar castes in the South on the Mandal commission's list. There is also a distinction between the caste 'taxonomy' in Jaffrelot's (2003) classifications of Indian politicians and the Mandal commission list with the former denoting the top layer among cultivating castes as 'intermediate'. Jaffrelot's 'intermediate group' includes among others the aforementioned *Jats*, plus *Reddy* and *Kamma* in Andhra Pradesh and *Vokkaligas* and *Lingayats* in Karnataka. Apart from us preferring 'upper' or 'forward' to 'intermediate' our classification is also for the main and most numerous groups (e.g *Khandayats* in Orissa, *Patidars* in Gujarat) consistent with Jaffrelot (2003).[6]

Table A1 provides a listing of upper castes based on our definition and begins with all India upper caste *jatis*; these are classified as upper castes in all states. The state listing provides additional upper caste *jatis*, which are either sub-groups of the main

[6] The state-wise official lists of STs, SCs and OBCs provide a rich source of information and were extensively consulted to cross check the SC and ST classifications in the raw data.

jatis (*Jats* or *Rajputs*, say) or belong to a different upper caste social group (e.g *Mahajan*; *Leva Patel*). Note that the following list is based exclusively on *jatis* that feature in the panel data set/village level social composition data. If a state is not specifically listed (e.g. Maharashtra), all upper caste groups in that state are already included in the ALL INDIA row. Notice also that the *jatis* in the ALL India row are by far the most numerous in the North. A careful reader may also notice that while Andhra castes and Kerala *Nayars* are included in Tamil Nadu, this is not the case the other way around. This is a co-incidence – there are no upper caste households from Tamil Nadu amongst our Andhra Pradesh panel households.

	Upper castes
ALL INDIA	Brahmin, Bhumihar, Rajput (general, Thakur), Kayastha, Kshatriya, Khatri, Maratha, Jat (Sikh and Hindu), Marwari, Bania (e.g. Agarwal, Gupta) (plus equivalents in the South: *Vysya* in Andhra Pradesh, *Chettiar* in Tamil Nadu)
ADDITIONAL BY STATE	
Himachal Pradesh	Rajput (Suniar), Choudhary
Punjab	Rajput (Suniar), Kamboj (Sikh), Choudhary, Mahant (Sikh), Arora, Ahluwalia, Mahajan, Sood, Visnoi
Uttaranchal	Rana
Haryana	Rajput (Chauhan, Bishnoi), Jat (Jhangi), Kamboj (Sikh)
Rajasthan	Choudhary, Mahajan
Gujarat	Patel (general, Patidar, Leva, Kadava), Rajput (Jadeja [Chandravanshi], Parmar, Solanki), Darbar
Uttar Pradesh	Rajput (Chauhan, Negi [Gharwali]), Srivastava, Choudhary
West Bengal	Pokhrel, Dahal, Chettri, Mahishya, Sadgop, Roy
Orissa	Patnaik (general, Karan), Pradhan, Khandayat, Odia, Kalandi
Madhya Pradesh	Jat (Tomar), Choudhary, Maharaj
Andhra Pradesh	Reddy, Kapu [Balija, Telaga], Kamma [Naidu], Velama, Chowdary, Rajulu
Karnataka	Lingayat, Vokkaliga
Tamil Nadu	Mudaliar, Vellalar, Nayar, Reddy, Naidu, Kamma Naidu
Kerala	Nayar (Nair)

References additional to those in the main text:

Chowdhry, P. (2007): Contentious Marriages, Eloping Couples: Gender, Caste and Patriarchy in Northern India, *Oxford University Press*.

Epstein, T. S., A.P. Suryanarayana and T. Thimmegowda (1998): Village Voices – 40 Years of Rural Transformation in South India, *Sage Publications.*

Jeffrey, C. (2001): A Fist is Stronger than Five Fingers: Caste and dominance in rural North-India, *Transactions of the Institute of British Geographers*, 25(2): 1-30.

Jeffrey, C., P. Jeffery and R. Jeffery (2007): Degrees without Freedom: Education,

Masculinities and Unemployment in North-India, *Stanford University Press*.
Singh, K. S. (1982): People of India – An Introduction, Anthropological Survey of India.
Srinivas, M. N. (1978): The remembered village, *Oxford University Press*.

Appendix 2

Table A2.1 Estimation results of the effects on income of social identity, village regime and demographic and locational controls

Model:	Social identity terms				Plus village regime				Plus agro-ecological zones				Plus state dummies				Plus demographic controls			
	Round 1		Round 2		Round 1		Round 2		Round 1		Round 2		Round 1		Round 2		Round 1		Round 2	
	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t
Social identity:																				
SC	-0.467	-16.32	-0.608	-18.43	-0.383	-8.21	-0.506	-10.15	-0.39	-8.49	-0.513	-10.27	-0.396	-8.53	-0.516	-10.25	-0.38	-8.58	-0.45	-9.45
ST	-0.461	-11.15	-0.602	-13.25	-0.316	-4.37	-0.461	-5.67	-0.302	-4.26	-0.422	-5.19	-0.302	-4.28	-0.441	-5.54	-0.315	-4.89	-0.373	-5.35
OBC	-0.271	-9.13	-0.372	-11.65	-0.296	-4.85	-0.31	-4.4	-0.276	-4.56	-0.297	-4.27	-0.272	-4.51	-0.289	-4.36	-0.231	-4.1	-0.241	-3.92
MUS	-0.41	-10.22	-0.53	-11.21	-0.294	-5.34	-0.445	-5.96	-0.275	-4.95	-0.438	-5.87	-0.283	-5.11	-0.465	-6.46	-0.208	-3.68	-0.323	-4.71
Village regime:																				
SC X DSC					0.037	0.49	0.265	2.74	0.103	1.37	0.257	2.7	0.117	1.57	0.281	2.92	0.145	2.04	0.299	3.06
ST X DST					-0.088	-1.16	-0.05	-0.59	-0.056	-0.77	-0.018	-0.22	-0.05	-0.69	0.002	0.02	-0.018	-0.27	0.003	0.05
OBC X DOBC					0.167	2.88	0.113	1.74	0.138	2.41	0.101	1.6	0.131	2.31	0.097	1.63	0.106	1.97	0.094	1.75
MUS X DMUS					-0.023	-0.29	0.164	1.68	-0.064	-0.75	0.086	0.85	-0.053	-0.6	0.103	1.06	-0.016	-0.2	0.124	1.37
UC X DUC					0.198	3.84	0.288	5.14	0.124	2.45	0.178	3.17	0.112	2.25	0.114	2.11	0.105	2.22	0.109	2.13
SC X DUC					0.088	2.04	0.139	3.01	0.021	0.45	0.042	0.89	0.013	0.29	-0.02	-0.43	0.026	0.6	-0.024	-0.54
ST X DUC					-0.02	-0.15	0.26	2.1	-0.105	-0.85	0.131	1.1	-0.095	-0.75	0.101	0.85	-0.023	-0.2	0.103	0.95
OBC X DUC					0.176	2.85	0.156	2.32	0.084	1.35	0.05	0.74	0.068	1.11	-0.014	-0.22	0.032	0.55	-0.011	-0.18
MUS X DUC					0.013	0.16	0.142	1.46	-0.086	-1.04	0.029	0.31	-0.068	-0.82	0.04	0.43	-0.037	-0.46	0.026	0.29
Agro-ecological zones:																				
aez2									0.272	2.47	0.273	3.32	0.634	1.67	0.484	1.8	0.449	1.37	0.546	2.07
aez3									0.595	8.79	-0.202	-1.44	0.599	1.76	-0.103	-0.39	0.439	1.51	-0.15	-0.58
aez4									0.114	1.72	0.113	1.7	0.382	1.03	0.343	1.3	0.172	0.54	0.393	1.51
aez5									0.078	0.93	-0.206	-2.44	0.389	1.08	0.175	0.64	0.165	0.54	0.178	0.66
aez6									0.317	4.63	-0.054	-0.78	0.606	1.74	0.282	1.2	0.467	1.58	0.238	1.02
aez7									0.48	3.45	-0.009	-0.07	0.315	0.81	-0.183	-0.62	0.198	0.58	-0.167	-0.57
aez8									0.211	2.86	0.059	0.66	0.121	0.36	0.293	1.11	-0.065	-0.23	0.194	0.75
aez9									0.279	3.47	0.129	1.75	0.503	1.35	0.421	1.52	0.27	0.84	0.412	1.53
aez10									-0.019	-0.23	-0.297	-4.46	0.214	0.62	0.304	1.23	0.021	0.07	0.289	1.19
aez11									0.036	0.48	-0.116	-1.64	0.262	0.82	0.555	2.61	0.14	0.53	0.532	2.62
aez12									-0.285	-3.94	-0.384	-4.55	-0.043	-0.15	0.193	1.07	-0.165	-0.73	0.15	0.85

Model:	Social identity terms				Plus village regime				Plus agro-ecological zones				Plus state dummies				Plus demographic controls			
	Round 1		Round 2		Round 1		Round 2		Round 1		Round 2		Round 1		Round 2		Round 1		Round 2	
	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t
aez13									-0.097	-0.94	-0.275	-2.42	0.07	0.17	0.268	0.89	-0.095	-0.27	0.283	0.98
aez14									-0.083	-1.06	0.229	2.96	0.338	0.86	0.296	1.05	0.066	0.2	0.278	1.01
aez17									-0.629	-2.33	-0.27	-1.1	-0.61	-2.25	-0.214	-0.9	-0.681	-2.42	-0.179	-0.83
aez18									-0.197	-2.73	0.374	5.26	-0.221	-0.65	0.446	1.94	-0.35	-1.21	0.418	1.86
aez19									-0.255	-1.73	-0.605	-4.41	0.032	0.1	0.121	0.54	-0.023	-0.09	0.059	0.27
aez20									0.253	1.94	0.176	1.18	0.23	0.66	0.167	0.66	0.02	0.07	0.081	0.32
State dummy variables:																				
Bihar													-0.187	-0.82	-0.469	-2.11	-0.034	-0.15	-0.323	-1.5
Gujarat													-0.337	-1.87	-0.138	-0.66	-0.242	-1.4	-0.132	-0.64
Haryana													-0.083	-0.46	0.045	0.22	0.108	0.62	0.086	0.43
Himachal Pradesh													-0.486	-2.12	0.026	0.11	-0.278	-1.26	0.056	0.25
Karnataka													-0.544	-3.46	-0.273	-1.81	-0.407	-2.67	-0.114	-0.7
Kerala													0.143	0.49	0.597	2.39	0.227	0.76	0.562	2.32
Madhya Pradesh													-0.249	-1.71	-0.535	-3	-0.097	-0.69	-0.466	-2.61
Maharashtra													-0.318	-3.72	-0.266	-1.94	-0.258	-3.14	-0.209	-1.52
Orissa													-0.36	-2.08	-0.666	-4.76	-0.347	-2.03	-0.581	-4.07
Punjab													-0.252	-1.3	0.142	0.68	-0.097	-0.51	0.167	0.81
Rajasthan													-0.504	-2.84	-0.244	-1.19	-0.319	-1.85	-0.165	-0.82
Tamil Nadu													0.064	0.9	-0.21	-1.34	0.096	1.32	-0.145	-0.95
Tripura													.	.	.	.	.	.	.	.
Uttar Pradesh													-0.382	-2.02	-0.44	-2.1	-0.2	-1.09	-0.329	-1.61
West Bengal													-0.041	-0.13	0.031	0.16	-0.118	-0.44	0.034	0.17
Uttaranchal													-0.442	-1.63	-0.228	-0.99	-0.23	-0.86	-0.129	-0.56
Chattisgarh													-0.264	-1.45	-0.629	-3.92	-0.225	-1.24	-0.567	-3.55
Jharkhand													-0.028	-0.14	-0.145	-0.76	0.046	0.24	-0.129	-0.72
Demographic controls																				
Sex of hh head (male = 1)																	0.03	1.46	-0.012	-0.53
# males aged 0-5																	-0.146	-12.63	-0.179	-13.83
# males aged 6-14																	-0.112	-11.94	-0.142	-15.13

Model:	Social identity terms				Plus village regime				Plus agro-ecological zones				Plus state dummies				Plus demographic controls			
	Round 1		Round 2		Round 1		Round 2		Round 1		Round 2		Round 1		Round 2		Round 1		Round 2	
	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t
# males aged 15-19																	-0.006	-0.45	-0.036	-2.63
# males, aged 20-24																	0.073	4.72	0.06	3.94
# males, aged 25-49																	0.104	5.76	0.108	6.64
# males, aged 50-59																	0.176	7.76	0.136	5.82
# males, aged 60 +																	0.074	3.42	0.056	2.7
# females, aged 0-5																	-0.13	-12.3	-0.154	-11.57
# females, aged 6-14																	-0.116	-13.64	-0.134	-13.65
# females, aged 15-19																	-0.065	-4.43	-0.089	-7.31
# females, aged 20-24																	-0.016	-0.8	-0.015	-0.84
# females, aged 25-49																	0.07	3.41	0.09	4.81
# females, aged 50-59																	0	0.01	0.099	3.88
# females, aged 60 +																	0.014	0.59	-0.031	-1.37
# of couples in household																	-0.021	-1.25	0.05	2.99
Constant	8.961	341.8	9.193	320.64	8.836	228.83	9.007	210.91	8.792	143.24	9.092	144.46	8.82	25.98	9.036	39.74	8.948	31.24	8.938	38.92
R squared	0.062		0.095		0.067		0.107		0.115		0.162		0.137		0.200		0.212		0.284	
N	9108		9108		9108		9108		9108		9108		9108		9108		9108		9108	

Source: HDPI-I ("round 1") and II ("round 2") surveys, panel households only; authors' calculations.
Notes: Dependent variable is the natural logarithm of annual per capita household income in constant 1993/94 prices, with round 2 figures converted using NSSO state-specific rural CPIs. Random effects, with standard errors that are robust to heteroskedasticity and clustering within villages; robust t-statistics are reported.

Table A2.2 Estimation results of the effects on income of social identity, village regime and additional controls: village infrastructure, household education and land

Controls added:	Village infrastructure				Plus education				Plus land			
	Round 1		Round 2		Round 1		Round 2		Round 1		Round 2	
	b	t	b	t	b	t	b	t	b	t	b	t
Social identity:												
SC	-0.378	-8.53	-0.447	-9.4	-0.255	-6.17	-0.316	-7.18	-0.166	-4.27	-0.248	-6.06
ST	-0.317	-4.92	-0.372	-5.34	-0.177	-2.85	-0.232	-3.49	-0.157	-2.73	-0.202	-3.08
OBC	-0.226	-3.95	-0.235	-3.83	-0.154	-2.85	-0.159	-2.83	-0.136	-2.81	-0.128	-2.62
MUS	-0.215	-3.86	-0.323	-4.67	-0.119	-2.17	-0.216	-3.31	-0.07	-1.38	-0.138	-2.22
Village regime:												
SC X DSC	0.135	1.95	0.285	2.92	0.129	2.01	0.249	2.7	0.067	1.09	0.202	2.41
ST X DST	0.006	0.09	0.024	0.35	-0.008	-0.12	0.009	0.14	0.012	0.21	-0.009	-0.14
OBC X DOBC	0.104	1.91	0.09	1.68	0.101	1.94	0.074	1.5	0.067	1.51	0.03	0.68
MUS X DMUS	0.003	0.03	0.149	1.65	0.005	0.07	0.152	1.8	-0.047	-0.74	0.088	1.15
UC X DUC	0.112	2.39	0.11	2.12	0.12	2.66	0.108	2.16	0.032	0.79	0.026	0.56
SC X DUC	0.026	0.6	-0.028	-0.64	0.015	0.38	-0.041	-0.98	0.044	1.18	-0.018	-0.46
ST X DUC	-0.025	-0.21	0.101	0.92	-0.006	-0.05	0.124	1.24	0.003	0.04	0.138	1.51
OBC X DUC	0.026	0.43	-0.018	-0.31	0.039	0.69	-0.01	-0.18	0.111	2.25	0.023	0.49
MUS X DUC	-0.034	-0.43	0.012	0.13	-0.004	-0.05	0.05	0.54	0.062	0.83	0.026	0.3
Agro-ecological zones:												
aez2	0.167	0.53	0.369	1.24	0.176	0.59	0.32	1.25	-0.021	-0.08	0.245	0.93
aez3	0.415	1.54	-0.291	-0.96	0.358	1.38	-0.31	-1.13	0.141	0.62	-0.33	-1.24
aez4	-0.101	-0.33	0.196	0.66	-0.139	-0.48	0.111	0.44	-0.189	-0.73	0.142	0.55
aez5	-0.073	-0.25	0.011	0.04	-0.055	-0.2	-0.049	-0.19	-0.169	-0.69	-0.043	-0.17
aez6	0.325	1.18	0.115	0.42	0.352	1.34	0.081	0.35	0.242	1.03	0.044	0.19
aez7	0.14	0.45	-0.321	-1.01	0.217	0.71	-0.276	-1.02	0.036	0.13	-0.302	-1.12
aez8	-0.04	-0.15	0.074	0.25	0.016	0.06	0.095	0.37	-0.127	-0.55	0.089	0.34
aez9	0.033	0.11	0.214	0.7	-0.002	-0.01	0.091	0.34	-0.041	-0.16	0.156	0.58
aez10	-0.149	-0.53	0.13	0.46	-0.206	-0.77	0.004	0.01	-0.256	-1.08	0.029	0.12
aez11	0.044	0.17	0.289	1.16	0.029	0.12	0.249	1.17	-0.114	-0.56	0.199	0.9
aez12	-0.123	-0.57	-0.001	-0.01	-0.074	-0.37	0.038	0.21	-0.176	-1.02	0.004	0.02

Controls added:	Village infrastructure				Plus education				Plus land			
	Round 1		Round 2		Round 1		Round 2		Round 1		Round 2	
	b	t	b	t	b	t	b	t	b	t	b	t
aez13	-0.359	-1.05	0.084	0.26	-0.289	-0.88	-0.029	-0.1	-0.143	-0.5	0.093	0.33
aez14	-0.135	-0.41	0.087	0.28	-0.287	-0.9	-0.094	-0.34	-0.171	-0.61	0.045	0.17
aez17	-0.543	-2.68	-0.133	-0.67	-0.511	-2.59	-0.15	-0.79	-0.408	-1.9	-0.106	-0.58
aez18	-0.304	-1.16	0.238	0.89	-0.263	-1.04	0.243	1.06	-0.306	-1.35	0.182	0.78
aez19	-0.025	-0.1	-0.069	-0.27	0.029	0.12	-0.026	-0.12	-0.069	-0.33	-0.001	0
aez20	-0.023	-0.08	-0.056	-0.19	-0.069	-0.26	-0.123	-0.49	-0.078	-0.32	-0.055	-0.22
State dummy variables:												
Bihar	0.168	0.72	-0.26	-1.19	0.163	0.72	-0.157	-0.8	-0.145	-0.74	-0.25	-1.33
Gujarat	-0.045	-0.25	-0.089	-0.43	0.022	0.12	0.012	0.07	0.003	0.02	-0.067	-0.4
Haryana	0.313	1.72	0.119	0.58	0.4	2.2	0.219	1.22	0.225	1.37	0.17	1.01
Himachal Pradesh	-0.035	-0.16	0.114	0.5	0.124	0.55	0.239	1.14	0.014	0.07	0.207	1.08
Karnataka	-0.342	-2.28	-0.106	-0.63	-0.284	-1.94	-0.09	-0.57	-0.342	-2.28	-0.104	-0.65
Kerala	0.242	0.82	0.499	1.89	0.25	0.89	0.435	1.81	0.254	0.93	0.439	1.94
Madhya Pradesh	0.184	1.19	-0.409	-2.23	0.292	1.86	-0.264	-1.7	0.099	0.7	-0.362	-2.45
Maharashtra	-0.167	-1.83	-0.22	-1.54	-0.143	-1.39	-0.215	-1.71	-0.183	-2.03	-0.226	-2.09
Orissa	-0.297	-1.93	-0.58	-4.01	-0.268	-1.66	-0.577	-4.48	-0.265	-1.82	-0.536	-4.37
Punjab	0.081	0.42	0.207	0.99	0.181	0.94	0.298	1.61	-0.059	-0.34	0.221	1.28
Rajasthan	-0.085	-0.47	-0.105	-0.52	0.023	0.13	0.016	0.09	-0.094	-0.58	-0.061	-0.37
Tamil Nadu	0.038	0.44	-0.165	-1.05	0.008	0.1	-0.206	-1.37	0.033	0.42	-0.214	-1.51
Tripura	.	.	.	.	.	.	.	.	.	.	.	.
Uttar Pradesh	0.024	0.13	-0.276	-1.33	0.122	0.64	-0.17	-0.93	-0.016	-0.09	-0.231	-1.34
West Bengal	-0.148	-0.61	-0.113	-0.47	-0.086	-0.37	-0.068	-0.34	-0.203	-0.98	-0.06	-0.29
Uttaranchal	0.018	0.07	-0.066	-0.28	0.188	0.75	0.116	0.53	-0.037	-0.17	0.005	0.02
Chattisgarh	-0.109	-0.65	-0.455	-2.67	-0.03	-0.17	-0.39	-2.56	-0.008	-0.05	-0.372	-2.55
Jharkhand	0.043	0.25	-0.083	-0.46	0.058	0.33	-0.086	-0.51	0.08	0.49	-0.005	-0.03
Demographic controls:												
Sex of hh head (male = 1)	0.03	1.46	-0.013	-0.57	0.037	1.87	-0.015	-0.67	0.038	2.09	-0.02	-0.92
# males aged 0-5	-0.145	-12.58	-0.178	-13.75	-0.138	-12.4	-0.16	-13.26	-0.149	-15.44	-0.15	-13.33
# males aged 6-14	-0.112	-11.91	-0.141	-15.03	-0.102	-11.58	-0.114	-12.47	-0.125	-15.59	-0.125	-14.69

Controls added:	Village infrastructure				Plus education				Plus land			
	Round 1		Round 2		Round 1		Round 2		Round 1		Round 2	
	b	t	b	t	b	t	b	t	b	t	b	t
# males aged 15-19	-0.006	-0.48	-0.035	-2.61	-0.047	-3.61	-0.073	-5.33	-0.077	-6.53	-0.094	-7.25
# males, aged 20-24	0.073	4.69	0.06	3.98	0.012	0.78	-0.014	-0.98	-0.01	-0.73	-0.032	-2.22
# males, aged 25-49	0.103	5.74	0.108	6.64	0.041	2.4	0.01	0.61	0.013	0.86	-0.011	-0.75
# males, aged 50-59	0.178	7.81	0.135	5.81	0.115	5.08	0.065	2.87	0.058	2.91	0.031	1.5
# males, aged 60 +	0.075	3.45	0.056	2.71	0.028	1.33	0.003	0.14	-0.031	-1.73	-0.037	-1.95
# females, aged 0-5	-0.131	-12.33	-0.153	-11.53	-0.123	-12.13	-0.135	-10.7	-0.132	-14.67	-0.137	-11.81
# females, aged 6-14	-0.116	-13.6	-0.134	-13.69	-0.111	-13.53	-0.116	-12.59	-0.128	-17.11	-0.13	-15.1
# females, aged 15-19	-0.066	-4.49	-0.089	-7.32	-0.13	-8.32	-0.135	-10.42	-0.134	-9.71	-0.145	-10.6
# females, aged 20-24	-0.018	-0.87	-0.017	-0.93	-0.092	-4.53	-0.102	-5.48	-0.102	-5.62	-0.111	-6.5
# females, aged 25-49	0.067	3.31	0.088	4.71	-0.005	-0.25	-0.007	-0.36	-0.036	-1.97	-0.045	-2.51
# females, aged 50-59	-0.002	-0.1	0.097	3.81	-0.055	-2.41	0.007	0.28	-0.075	-3.9	-0.021	-0.85
# females, aged 60 +	0.011	0.47	-0.033	-1.49	-0.042	-1.86	-0.104	-4.7	-0.077	-3.85	-0.13	-6.25
# of couples in household	-0.019	-1.14	0.051	3.07	0.026	1.6	0.106	6.49	0.007	0.46	0.067	4.47
Village infrastructure:												
Ln(village population)	0.018	0.96	-0.006	-0.38	0.007	0.4	-0.012	-0.79	0.012	0.74	-0.009	-0.6
School access:												
Primary	0.008	0.08	0.363	1.55	-0.015	-0.13	0.274	2.35	-0.009	-0.1	0.305	2
Middle	-0.196	-2.95	-0.126	-1.81	-0.185	-2.8	-0.05	-0.73	-0.125	-2.14	-0.064	-0.98
Lower secondary	0.045	0.73	-0.095	-1.56	0.059	0.97	-0.032	-0.55	0.064	1.13	-0.077	-1.35
Higher secondary	-0.012	-0.3	-0.072	-1.32	-0.005	-0.12	-0.041	-0.78	-0.023	-0.64	-0.056	-1.07
Graduate	-0.019	-0.52	-0.068	-1.32	-0.021	-0.59	-0.036	-0.71	-0.02	-0.64	-0.048	-0.97
Vocational	-0.017	-0.39	-0.025	-0.4	-0.043	-1.03	-0.011	-0.18	-0.002	-0.06	-0.018	-0.29
Medical access:												
Doctor	0	-0.23	-0.003	-1.56	0	-0.06	-0.002	-1.31	-0.001	-0.78	-0.002	-1.42
Clinic	-0.019	-0.58	-0.03	-0.95	-0.039	-1.19	-0.032	-1.04	-0.029	-1.01	-0.013	-0.45
Road access:												
Feeder	0.07	1.35	-0.013	-0.22	0.065	1.28	-0.019	-0.33	0.04	0.83	-0.018	-0.33
Tarmac	0.114	2.05	0.008	0.13	0.099	1.81	-0.004	-0.07	0.077	1.52	-0.008	-0.15
Bus stop	0.018	0.48	0.002	0.08	0.016	0.44	0.009	0.32	0.012	0.36	0.012	0.44

Controls added:	Village infrastructure				Plus education				Plus land			
	Round 1		Round 2		Round 1		Round 2		Round 1		Round 2	
	b	t	b	t	b	t	b	t	b	t	b	t
Railway station	0.106	2.43	0.103	2.51	0.07	1.52	0.066	1.61	0.074	1.9	0.055	1.43
Post office	0.112	3.16	0.027	0.82	0.108	3.1	0.013	0.42	0.11	3.46	0.015	0.52
Bank/credit market	-0.048	-1.39	-0.012	-0.34	-0.055	-1.65	-0.048	-1.41	-0.027	-0.89	-0.022	-0.69
Market/mandi	-0.068	-2.01	0.002	0.06	-0.077	-2.28	0.005	0.17	-0.064	-2.1	0.013	0.48
Max. educational achievement in the household (of those 15+):												
Males												
Up to primary					0.08	4.23	0.035	1.58	0.066	3.95	0.042	1.99
Middle					0.177	8.14	0.129	6.1	0.143	7.46	0.13	6.59
Matriculation					0.272	9.5	0.31	10.9	0.206	8.21	0.287	10.43
Higher secondary					0.341	9.97	0.365	11.67	0.278	9.1	0.312	10.08
Graduate and above					0.582	13.88	0.608	15.76	0.455	11.84	0.513	14.08
Females												
Up to primary					0.083	3.76	0.087	3.98	0.061	3.23	0.071	3.45
Middle					0.191	6.52	0.116	5.15	0.138	5.27	0.087	4.15
Matriculation					0.246	6.39	0.149	4.88	0.162	4.59	0.14	4.91
Higher secondary					0.192	2.98	0.329	8.05	0.212	3.43	0.262	6.58
Graduate and above					0.264	2.89	0.335	6.47	0.289	3.55	0.286	5.92
Household land:												
Land owned in acres									0.022	5.37	0.029	6.85
Land gross irrigated in acres									0.047	10.36	0.036	4.78
Constant	8.75	28.5	9.203	28.92	8.655	28.99	9.073	32.86	8.811	33.22	9.154	33.12
R squared	0.225		0.288		0.270		0.341		0.426		0.418	
N	9108		9108		9108		9108		9108		9108	

Source and Notes: as for Table A2.1

Table A2.3 Estimation results of the effects on income of social identity, village regime and additional controls: by social group

Model:	SC				ST				OBC				MUS				UC			
	Round 1		Round 2		Round 1		Round 2		Round 1		Round 2		Round 1		Round 2		Round 1		Round 2	
	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t
Village regime:																				
DSC	0.025	0.29	0.3	2.93																
DST					-0.01	-0.13	0.022	0.31												
DOBC									0.11	1.84	0.079	1.71								
DMUS													0.07	0.7	0.151	1.3				
DUC	-0.011	-0.26	0.025	0.55	-0.067	-0.51	0.111	0.87	0	0	-0.062	-1.22	0.098	1.03	0.105	0.91	0.074	1.42	0.059	1.06
Agro-ecological zones:																				
aez2	-0.296	-0.66	-0.711	-1.62	1.318	0.63	-1.711	-0.68	-0.087	-0.14	1.027	1.38	2.044	1.9	0.545	0.44	0.702	0.64	0.366	0.33
aez3	-0.367	-1.04	-1.044	-2.7	1.379	0.68	-2.176	-0.81	0.086	0.14	0.57	0.76	1.395	2.14	-0.509	-0.7	1.432	1.87	0.835	1.05
aez4	-0.552	-1.24	-0.71	-1.62	0.975	0.48	-2.409	-0.97	-0.387	-0.62	0.698	0.94	1.445	1.39	1.05	0.88	0.559	0.51	0.314	0.28
aez5	-0.506	-1.17	-0.792	-1.81	1.048	0.52	-2.058	-0.84	-0.385	-0.62	0.813	1.07	1.503	1.42	0.857	0.8	0.296	0.27	-0.366	-0.32
aez6	-0.462	-1.22	-0.797	-2.13	1.015	0.51	-2.077	-0.86	0.049	0.08	0.83	1.14	1.869	2.37	0.811	0.81	1.956	2.39	1.002	1.22
aez7	-0.741	-1.71	-1.342	-2.8	0.142	0.07	-1.667	-0.58	-0.296	-0.46	0.51	0.66	1.07	1.53	0.113	0.15	1.865	2.3	0.858	1.14
aez8	-0.879	-2.52	-0.461	-1.21	1.099	0.54	-1.632	-0.61	-0.53	-0.85	0.507	0.66	1.195	2.08	1.268	1.7	1.247	1.63	1.148	1.59
aez9	-0.454	-1.04	-0.632	-1.43	3.137	1.29	5.653	2.14	-0.279	-0.44	0.764	1.02	1.563	1.44	1.012	0.83	0.504	0.46	0.219	0.19
aez10	-0.763	-1.86	-0.587	-1.4	0.808	0.41	-2.222	-0.91	-0.562	-0.93	0.695	0.95	2.089	2.03	1.226	1.09	0.416	0.39	-0.048	-0.04
aez11	-0.78	-2.37	0.139	0.43	1.122	0.57	-1.964	-0.82	-0.372	-0.64	0.804	1.14	0.32	0.46	0.335	0.56	0.113	0.11	0.533	0.35
aez12	-0.93	-3.18	-0.294	-1.11	0.547	0.28	-2.497	-1.05	-0.406	-0.74	0.405	0.58	0.196	0.51	0.148	0.36	1.024	1.52	1.417	2.11
aez13	-0.52	-1.04	-0.958	-2.05	0	.	0	.	-0.833	-1.21	0.642	0.84	1.121	1.03	0.921	0.74	0.134	0.12	-0.002	0
aez14	-0.69	-1.45	-0.56	-1.24	0.658	0.32	-1.692	-0.63	-0.516	-0.81	0.621	0.82	2.22	2.1	1.058	0.86	0.448	0.4	0.028	0.02
aez17	-0.422	-2.78	-0.129	-0.75	-1.718	-2.06	-0.124	-0.11	-0.491	-1.2	-0.303	-0.64	0.001	0	0.509	1.92	-1.092	-2.64	-0.301	-0.69
aez18	-1.073	-3.12	-0.293	-0.82	0	.	0	.	-0.758	-1.25	1.12	1.52	-0.098	-0.12	1.031	1.46	0.823	1.08	0.786	1.07
aez19	-0.826	-2.6	-0.239	-0.77	0.565	0.23	-3.907	-1.23	-0.194	-0.3	0.574	0.79	-0.053	-0.11	-0.045	-0.1	0.923	1.35	0.735	1.05
aez20	-0.814	-1.97	-0.232	-0.56	-3.358	-0.46	-10.469	-1.43	-0.551	-0.9	0.45	0.6	1.394	1.74	1.449	1.5	1.756	2.16	0.772	0.94
State dummy variables:																				
Bihar	-0.417	-1.15	0.057	0.17	0	.	0	.	0.185	0.52	-0.128	-0.44	0.056	0.06	-0.809	-0.96	1.194	1.48	0.81	0.92
Gujarat	-0.33	-1.13	-0.036	-0.11	-0.66	-1.58	-0.277	-0.34	-0.055	-0.23	0.019	0.07	-0.615	-0.73	-0.713	-0.86	0.924	1.16	1.252	1.47

Model:	SC				ST				OBC				MUS				UC			
	Round 1		Round 2		Round 1		Round 2		Round 1		Round 2		Round 1		Round 2		Round 1		Round 2	
	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t
Haryana	-0.129	-0.46	0.161	0.5	0	.	0	.	0.343	1.53	0.39	1.5	-0.258	-0.31	-0.536	-0.64	1.182	1.48	1.201	1.41
Himachal Pradesh	-0.263	-0.8	0.109	0.32	0	.	0	.	0.11	0.36	0.585	1.75	-0.94	-1.05	-0.234	-0.26	0.655	0.8	1.136	1.29
Karnataka	-0.6	-1.89	0.305	1.71	0.152	0.56	-0.295	-0.37	-0.651	-1.28	0.031	0.14	-0.344	-1.03	-0.451	-1.11	-0.608	-1.79	-0.335	-0.75
Kerala	0.4	1.4	0.209	0.37	0	.	0	.	0.202	0.4	0.448	1.58	-0.179	-0.26	-0.278	-0.46	-0.343	-0.78	0.527	1.24
Madhya Pradesh	-0.28	-1.22	-0.499	-1.72	-0.295	-0.86	-0.247	-0.31	-0.014	-0.07	-0.233	-1.03	-0.737	-0.93	-1.168	-1.67	1.039	1.34	0.804	0.97
Maharashtra	-0.264	-1.8	0.047	0.26	-0.325	-1.26	-0.078	-0.1	-0.171	-1.32	-0.073	-0.42	-0.696	-2.06	-0.48	-1.2	-0.323	-1.03	-0.121	-0.28
Orissa	-0.399	-2.12	-0.974	-4.02	-0.577	-1.76	-0.206	-0.25	-0.542	-2.25	-0.281	-1.27	1.436	3.18	0.253	0.45	-0.088	-0.27	-0.728	-3.05
Punjab	-0.228	-0.79	0.29	0.9	0	.	0	.	-0.011	-0.05	0.324	1.19	-0.856	-0.96	-0.928	-1.11	1.06	1.32	1.333	1.56
Rajasthan	-0.369	-1.36	0.139	0.44	-0.573	-1.13	0.222	0.25	-0.191	-0.86	0.123	0.48	-0.459	-0.54	-0.349	-0.41	0.611	0.77	0.771	0.91
Tamil Nadu	0.062	0.56	-0.251	-1.52	-1.797	-5.66	-0.929	-4.95	0.107	0.63	0.108	0.4	-0.294	-1.52	-0.761	-2.24	0.08	0.36	-0.488	-2.14
Tripura	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Uttar Pradesh	-0.436	-1.4	-0.134	-0.42	0.509	1.1	0.656	0.75	-0.209	-0.84	-0.072	-0.28	-0.425	-0.49	-0.71	-0.87	0.97	1.22	0.566	0.66
West Bengal	-0.844	-2.76	-0.683	-2.12	1.03	0.51	-2.271	-0.85	-0.443	-0.78	0.624	0.92	0.904	1.62	0.13	0.19	1.112	1.55	0.775	1.17
Uttaranchal	-0.362	-1.05	-0.202	-0.59	0	.	0	.	-0.013	-0.04	0.299	0.85	-0.292	-0.34	-0.069	-0.08	0.803	0.92	0.916	0.99
Chattisgarh	-0.055	-0.23	-0.893	-3.12	-0.528	-1.59	-0.236	-0.28	-0.243	-1	-0.282	-1.29	0.761	1.06	-0.535	-0.62	0.98	1.33	0.038	0.03
Jharkhand	0.39	1.91	-0.88	-2.59	-0.34	-1.02	0.096	0.11	-0.252	-0.9	0.114	0.37	0.867	1.73	0.118	0.19	0.648	1.65	-0.169	-0.66
Demographic controls	0.024	0.62	0.073	1.67	0.026	0.4	-0.043	-0.57	0.042	1.09	-0.002	-0.05	0.061	1.04	-0.021	-0.23	0.022	0.48	-0.087	-1.93
Sex of hh head (male = 1)																				
# males aged 0-5	-0.155	-7.35	-0.143	-5.47	-0.184	-5.61	-0.263	-5.89	-0.153	-7.34	-0.171	-7.64	-0.135	-4.39	-0.146	-3.68	-0.132	-5.64	-0.183	-7.29
# males aged 6-14	-0.119	-7.22	-0.122	-6.31	-0.114	-4.31	-0.185	-7.68	-0.111	-6.55	-0.152	-9	-0.136	-4.42	-0.079	-3.43	-0.1	-5.06	-0.145	-7.17
# males aged 15-19	0.01	0.36	-0.043	-1.83	0.008	0.21	-0.071	-2.01	-0.003	-0.15	0.014	0.56	0.026	0.46	-0.084	-2.23	-0.028	-1.16	-0.046	-1.6
# males, aged 20-24	0.112	4.02	0.092	3.46	0.086	1.62	0.047	0.87	0.052	1.97	0.044	1.44	0.032	0.76	0.084	1.4	0.049	1.55	0.048	1.48
# males, aged 25-49	0.145	4.06	0.148	4.75	0.05	0.9	0.057	1	0.121	3.74	0.139	5.11	0.012	0.25	0.164	3.04	0.067	1.8	0.025	0.89
# males, aged 50-59	0.214	5.05	0.18	4.33	0.085	1.05	0.087	1.19	0.193	4.87	0.176	4.51	0.026	0.29	0.089	0.95	0.156	3.15	0.037	0.86
# males, aged 60 +	0.078	2.07	-0.009	-0.22	0.086	1.08	0.066	0.91	0.128	3.47	0.082	2.29	-0.06	-0.75	0.197	2.63	0.038	0.92	0.012	0.32
# females, aged 0-5	-0.12	-6.45	-0.149	-6.19	-0.158	-4.59	-0.165	-5.44	-0.125	-6.33	-0.153	-7.19	-0.12	-4.11	-0.175	-3.36	-0.148	-6.91	-0.131	-4.65
# females, aged 6-14	-0.132	-7.94	-0.155	-9.65	-0.11	-3.45	-0.157	-6.46	-0.106	-6.73	-0.131	-7.89	-0.101	-4.34	-0.085	-2.4	-0.119	-6.62	-0.123	-6.46
# females, aged 15-19	-0.023	-0.85	-0.057	-2.46	-0.139	-2.48	-0.068	-1.82	-0.067	-2.36	-0.101	-3.99	-0.096	-1.77	-0.109	-3.05	-0.037	-1.28	-0.082	-3.07
# females, aged 20-24	-0.014	-0.34	-0.012	-0.3	-0.103	-1.54	-0.032	-0.53	-0.045	-1.26	0.013	0.4	-0.001	-0.01	-0.08	-1.36	0.041	1.13	-0.021	-0.58

Model:	SC				ST				OBC				MUS				UC			
	Round 1		Round 2		Round 1		Round 2		Round 1		Round 2		Round 1		Round 2		Round 1		Round 2	
	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t	b	t
# females, aged 25-49	0.063	1.52	-0.025	-0.58	0.006	0.1	0	0	0.066	1.84	0.131	4.16	0.031	0.36	0.017	0.21	0.091	2.14	0.133	3.49
# females, aged 50-59	-0.066	-1.52	0.011	0.23	-0.012	-0.15	0.041	0.61	0.002	0.04	0.123	2.63	0.056	0.74	0.098	0.88	0.067	1.45	0.127	2.41
# females, aged 60 +	-0.062	-1.29	0.021	0.48	-0.019	-0.23	-0.077	-1.13	-0.011	-0.24	-0.015	-0.32	0.023	0.29	-0.131	-1.52	0.079	1.77	-0.055	-1.2
# of couples in household	-0.043	-1.29	-0.001	-0.03	0.033	0.56	0.068	1.27	-0.029	-1.02	0.021	0.68	0.027	0.5	0.084	1.89	-0.027	-0.73	0.09	2.41
Constant	9.46	27.36	9.35	25.71	8.182	4.14	11.043	4.13	9.131	14.9	8.027	10.86	7.622	13.74	8.323	11.94	7.594	10.09	8.189	11.49
R squared	0.180		0.267		0.225		0.328		0.176		0.226		0.229		0.228		0.196		0.226	
N	2252		2252		835		835		2909		2909		699		699		2413		2413	

Source and Notes: as for Table A2.1

Table A2.4 Estimation results of the effects on change in land/education of social identity, village regime and additional controls

Model:	Change in land owned		Change in highest level of female education (probit)		Change in highest level of male education (probit)	
	b	t	b	z	b	z
Social identity:						
SC	0.237	0.73	-0.875	-9.43	-0.704	-7.61
ST	0.548	0.97	-0.913	-7.59	-0.771	-5.96
OBC	-1.055	-1.72	-0.373	-3.17	-0.327	-3
MUS	-0.63	-1.18	-0.667	-4.61	-0.575	-4.08
Village regime:						
SC X DSC	-0.176	-0.81	0.187	1.09	0.198	1.29
ST X DST	-0.084	-0.15	-0.142	-1.16	-0.241	-1.79
OBC X DOBC	0.966	1.78	-0.004	-0.04	0.016	0.17
MUS X DMUS	0.327	0.63	-0.385	-1.95	-0.228	-1.16
UC X DUC	-0.186	-0.51	-0.096	-1	-0.099	-1.01
SC X DUC	-0.321	-0.94	0.092	1.1	-0.014	-0.16
ST X DUC	-0.369	-0.61	-0.176	-0.69	-0.304	-1.49
OBC X DUC	1.042	1.81	-0.14	-1.22	-0.244	-2.16
MUS X DUC	1.142	2.3	-0.37	-1.29	-0.207	-1.16
Agro-ecological zones:						
aez2	-3.636	-2.04	0.093	0.13	-0.177	-0.27
aez3	-1.402	-1.13	-0.498	-0.73	-0.457	-0.77
aez4	-2.253	-1.37	0.172	0.25	0.009	0.01
aez5	-3.22	-1.98	-0.016	-0.02	-0.582	-0.89
aez6	-1.52	-1.37	-0.421	-0.61	-0.764	-1.2
aez7	-1.83	-1.58	-0.656	-0.85	-1.151	-1.5
aez8	-1.494	-1.48	-0.419	-0.63	-0.547	-0.93
aez9	-3.284	-1.93	0.611	0.86	0.192	0.28
aez10	-1.448	-0.98	0.675	0.99	-0.089	-0.14
aez11	-1.784	-0.89	0.577	0.88	0.441	0.76
aez12	0.276	0.42	-0.589	-0.98	-0.59	-1.2
aez13	-2.601	-1.46	0.417	0.54	-0.165	-0.21
aez14	-3.38	-2.03	0.982	1.34	0.28	0.38
aez17	0.553	2.19	-0.387	-0.99	0.127	0.39

aez18	1.006	0.94	0.348	0.52	0.01	0.02
aez19	-0.984	-1.01	-0.234	-0.38	-0.066	-0.12
aez20	-1.888	-1.7	0.342	0.48	-0.677	-1.07
State dummy variables:						
Bihar	0.892	0.63	-0.955	-2.04	-0.559	-1.03
Gujarat	1.393	1.08	-0.555	-1.48	0.069	0.17
Haryana	1.232	0.98	-0.733	-1.97	-0.34	-0.82
Himachal Pradesh	1.625	1.24	-1.019	-2.35	-0.247	-0.46
Karnataka	2.098	1.67	0.496	1.67	0.411	1.07
Kerala	0.656	0.94	0.867	1.92	1.901	4
Madhya Pradesh	-0.451	-0.42	-1.253	-3.79	-0.354	-0.94
Maharashtra	-0.191	-0.27	0.325	1.26	0.541	1.88
Orissa	-1.697	-2.58	0.096	0.36	-0.014	-0.05
Punjab	1.809	1.41	-0.566	-1.5	-0.504	-1.2
Rajasthan	0.427	0.34	-1.387	-3.71	-0.588	-1.44
Tamil Nadu	0.675	1.62	0.275	1.5	0.116	0.73
Tripura	.	.	.	.	.	.
Uttar Pradesh	1.704	1.33	-1.072	-2.9	-0.739	-1.79
West Bengal	-1.323	-1.47	-0.142	-0.24	-0.611	-1.13
Uttaranchal	2.158	1.63	-1.309	-3.17	-0.504	-0.98
Chattisgarh	0.25	0.18	-1.028	-3.18	-0.916	-2.35
Jharkhand	-2.884	-2.85	-0.485	-1.62	-0.314	-0.94
Demographic controls						
Sex of hh head (male = 1)	0.238	1.27	0.038	0.82	-0.06	-1.18
# males aged 0-5	0.058	0.52	-0.154	-5.09	-0.111	-3.3
# males aged 6-14	0.119	1.32	-0.128	-5.71	-0.104	-4.59
# males aged 15-19	0.011	0.08	-0.143	-5.05	0.661	15.96
# males, aged 20-24	0.179	0.91	-0.021	-0.62	0.673	13.04
# males, aged 25-49	0.079	0.59	0.259	7.59	0.865	17.31
# males, aged 50-59	0.404	1.48	0.111	2.33	0.625	10.68
# males, aged 60 +	0.153	0.89	0.209	4.94	0.449	9.71
# females, aged 0-5	0.199	1.69	-0.153	-5.41	-0.109	-3.22
# females, aged 6-14	-0.082	-0.85	-0.083	-4.24	-0.031	-1.47
# females, aged 15-19	0.074	0.5	0.794	19.24	0.017	0.6
# females, aged 20-24	0.095	0.43	0.89	17.99	0.222	5.07
# females, aged 25-49	0.466	2.53	0.887	19.04	0.399	8.74

# females, aged 50-59	0.318	1.5	0.601	10.64	0.334	5.69
# females, aged 60 +	0.271	1.33	0.446	8.8	0.272	5.19
# of couples in household	-0.007	-0.03	-0.328	-8.25	-0.212	-5.44
Constant	-0.32	-0.29	-0.2	-0.31	0.303	0.51
(Pseudo) R squared	0.0211		0.2655		0.2464	
N	9108		9108		9108	

Source and Notes: as for Table A2.1

Appendix 3 Derivation of equation (4) and interpretation

The household group, enclave, oppression and proximity effects are not all separately identified and to facilitate the interpretation of our empirical results, this appendix show what is identified and how to interpret the estimated coefficients.

The parameter notation is as follows. The α's are household group effects, the β's are enclave effects, the γ's are oppression effects and the δ's are proximity effects. Oppression effects and proximity effects are with respect to UC dominated villages and, therefore, for UC households, by definition, equal to zero, i.e. we set $\gamma_{5t} = 0$ and $\delta_{5t} = 0$ (see also main text). To start with, we include all possible effects in the income equation.

$$\begin{aligned}\ln(Y_{ht}) = {} & \alpha_{0t} + \alpha_{1t}SC_h + \alpha_{2t}ST_h + \alpha_{3t}MUS_h + \alpha_{4t}OBC_h + \alpha_{5t}UC_h \\ & + \beta_{1t}SC_h \times DSC_{v(h)} + \beta_{2t}ST_h \times DST_{v(h)} + \beta_{3t}MUS_h \times DMUS_{v(h)} \\ & + \beta_{4t}OBC_h \times DOBC_{v(h)} + \beta_{5t}UC_h \times DUC_{v(h)} \\ & + \gamma_{1t}SC_h \times DUC_{v(h)} + \gamma_{2t}ST_h \times DUC_{v(h)} \\ & + \gamma_{3t}MUS_h \times DUC_{v(h)} + \gamma_{4t}OBC_h \times DUC_{v(h)} \\ & + \delta_{1t}SC_h \times DUC_{v(h)} + \delta_{2t}ST_h \times DUC_{v(h)} \\ & + \delta_{3t}MUS_h \times DUC_{v(h)} + \delta_{4t}OBC_h \times DUC_{v(h)} \\ & + \pi_t X_{ht} + \theta_h + \eta_{v(h)} + \varepsilon_{ht} \qquad \text{(A1)}\end{aligned}$$

As in the main text other explanatory variables are denoted by X_{ht} and the error terms by $\theta_h, \eta_{v(h)}$ and ε_{ht}. Next, we simplify the rows of eq.(A1) that correspond to the oppression and proximity effects as these cannot, at the outset, be separately identified for all household groups (as pointed out by the referee and the Associate Editor, only the net effects can).

$$\begin{aligned}\ln(Y_{ht}) = {} & \alpha_{0t} + \alpha_{1t}SC_h + \alpha_{2t}ST_h + \alpha_{3t}MUS_h + \alpha_{4t}OBC_h + \alpha_{5t}UC_h \\ & + \beta_{1t}SC_h \times DSC_{v(h)} + \beta_{2t}ST_h \times DST_{v(h)} + \beta_{3t}MUS_h \times DMUS_{v(h)} \\ & + \beta_{4t}OBC_h \times DOBC_{v(h)} + \beta_{5t}UC_h \times DUC_{v(h)} \\ & + (\gamma_{1t} + \delta_{1t})SC_h \times DUC_{v(h)} + (\gamma_{2t} + \delta_{2t})ST_h \times DUC_{v(h)} \\ & + (\gamma_{3t} + \delta_{3t})MUS_h \times DUC_{v(h)} + (\gamma_{4t} + \delta_{4t})OBC_h \times DUC_{v(h)} \\ & + \pi_t X_{ht} + \theta_h + \eta_{v(h)} + \varepsilon_{ht} \qquad \text{(A2)}\end{aligned}$$

Finally, since $SC_h + ST_h + MUS_h + OBC_h + UC_h = 1$, we substitute $UC_h = 1 - (SC_h + ST_h + MUS_h + OBC_h)$ in eq.(A2) and, after rearranging terms, this yields

$$
\begin{aligned}
\ln(Y_{ht}) = {} & (\alpha_{0t} + \alpha_{5t}) + (\alpha_{1t} - \alpha_{5t})SC_h + (\alpha_{2t} - \alpha_{5t})ST_h \\
& + (\alpha_{3t} - \alpha_{5t})MUS_h + (\alpha_{4t} - \alpha_{5t})OBC_h \\
& + \beta_{1t}SC_h \times DSC_{v(h)} + \beta_{2t}ST_h \times DST_{v(h)} \\
& + \beta_{3t}MUS_h \times DMUS_{v(h)} + \beta_{4t}OBC_h \times DOBC_{v(h)} \\
& + \beta_{5t} \times DUC_{v(h)} \\
& + (\gamma_{1t} + \delta_{1t} - \beta_{5t})SC_h \times DUC_{v(h)} + (\gamma_{2t} + \delta_{2t} - \beta_{5t})ST_h \times DUC_{v(h)} \\
& + (\gamma_{3t} + \delta_{3t} - \beta_{5t})MUS_h \times DUC_{v(h)} + (\gamma_{4t} + \delta_{4t} - \beta_{5t})OBC_h \times DUC_{v(h)} \\
& + \pi_t X_{ht} + \theta_h + \eta_{v(h)} + \varepsilon_{ht}.
\end{aligned}
\tag{A3}
$$

This is the specification in the previous version of our paper which we use in the simulations of the order of magnitude of the enclave, proximity and oppression effects in section IV.C: the interpretation of the coefficients corresponding with the DUC interactions includes enclave, oppression and proximity effects. In the previous version of the paper, we treated proximity effects as undifferentiated across social groups so that $\delta_{1t}=\delta_{2t}\ldots=\delta_{4t}$. This is a plausible assumption. Substituting in A3, the parenthesis in front of each interaction term then simplifies to $(\gamma_{it} + \delta_{1t} - \beta_5)$ with the last two terms now identical for all social groups. In our section IV.C simulations, we set the proximity gain to marginalized groups equal to the upper caste enclave effect, β_5 which, by assuming this is an upper bound on the proximity effect for the other social groups, allows for separate interpretations of the proximity and oppression effects.

To obtain a clearer interpretation of the oppression/proximity effects, we substitute $SC_h + ST_h + MUS_h + OBC_h + UC_h = 1$ in the fifth line of eq.(A3) so that $\beta_{5t} \times DUC_{v(h)}$ becomes $\beta_{5t} \times DUC_{v(h)} \left(SC_h + ST_h + MUS_h + OBC_h + UC_h\right)$. This yields

$$
\begin{aligned}
\ln(Y_{ht}) = {} & (\alpha_{0t} + \alpha_{5t}) + (\alpha_{1t} - \alpha_{5t})SC_h + (\alpha_{2t} - \alpha_{5t})ST_h \\
& + (\alpha_{3t} - \alpha_{5t})MUS_h + (\alpha_{4t} - \alpha_{5t})OBC_h \\
& + \beta_{1t}SC_h \times DSC_{v(h)} + \beta_{2t}ST_h \times DST_{v(h)} \\
& + \beta_{3t}MUS_h \times DMUS_{v(h)} + \beta_{4t}OBC_h \times DOBC_{v(h)} \\
& + \beta_{5t}UC_h \times DUC_{v(h)} \\
& + (\gamma_{1t} + \delta_{1t})SC_h \times DUC_{v(h)} + (\gamma_{2t} + \delta_{2t})ST_h \times DUC_{v(h)} \\
& + (\gamma_{3t} + \delta_{3t})MUS_h \times DUC_{v(h)} + (\gamma_{4t} + \delta_{4t})OBC_h \times DUC_{v(h)} \\
& + \pi_t X_{ht} + \theta_h + \eta_{v(h)} + \varepsilon_{ht}
\end{aligned}
\tag{A4}
$$

Eq.(A4) shows that the household group effects (α's) are relative to UC household and that the enclave effect (β's) are relative to non-own group and non-UC dominated villages. We discuss interpretations below. For non-UC households only the sum of the proximity and oppression effects is identified.

Interpretation of the estimated coefficients

Equation (A4) is equation 4 in the main text with the only difference being a different notations for the coefficients (e.g., $(\alpha_{1t} - \alpha_{5t})$ in equation (A4) is α_{1t} in equation 4) for convenience. Note that UC is the reference household.

The coefficients corresponding to the household groups need to be interpreted taking into account that there are enclave and oppression/proximity effects. For example, consider an SC household, hence SC_h=1 and the dummy variables for the other household groups are zero, i.e. $UC_h = 0, ST_h = 0, MUS_h = 0, OBC_h = 0$. The income equation for an SC household is given by

$$\ln(Y_{ht}) = (\alpha_{0t} + \alpha_{5t}) + (\alpha_{1t} - \alpha_{5t}) + \beta_{1t} DSC_{v(h)} + (\gamma_{1t} + \delta_{1t}) DUC_{v(h)} \qquad \text{(A5)}$$
$$+ \pi_t X_{ht} + \theta_h + \eta_{v(h)} + \varepsilon_{ht}.$$

The coefficient $(\alpha_{1t} - \alpha_{5t})$ corresponds to the effect of SC_h on income and is interpreted as a $(\alpha_{1t} - \alpha_{5t})$ x100 percentage difference in income between an SC and an UC household living in a village that is dominated neither by SC nor by UC households, i.e. when $DSC_{v(h)}$ and $DUC_{v(h)}$ are equal to zero, (ceteris paribus). The coefficient β_{1t} x100 is the percentage difference in income between *SC* households living in a village dominated by their own group, i.e. $DSC_{v(h)}$=1 and $DUC_{v(h)}$=0, and *SC* households living in a village dominated neither by SC nor by UC (again, the reference requires that $DSC_{v(h)}$ and $DUC_{v(h)}$ are both equal to zero). Likewise, the coefficient $(\gamma_{1t} + \delta_{1t})$ x100 is the percentage difference in income of *SC* households living in a *UC* dominated village and households living in a village dominated neither by SC nor by UC. We refer to the latter as the net effect of oppression and proximity for an SC household. The interpretations of the coefficients corresponding to ST, MUS or OBC households are similar.

Concerning UC households, we set $UC_h = 1, SC_h = 0 ST_h = 0, MUS_h = 0, OBC_h = 0$ in equation (A4), which yields

$$\ln(Y_{ht}) = (\alpha_{0t} + \alpha_{5t}) + \beta_{5t} DUC_{v(h)} + \pi_t X_{ht} + \theta_h + \eta_{v(h)} + \varepsilon_{ht}. \qquad \text{(A6)}$$

β_{5t} x100 is the difference in income between an UC household living in an UC dominated village compared to an UC household living in a non-UC dominated village (ceteris paribus).

Figure
Click here to download high resolution image

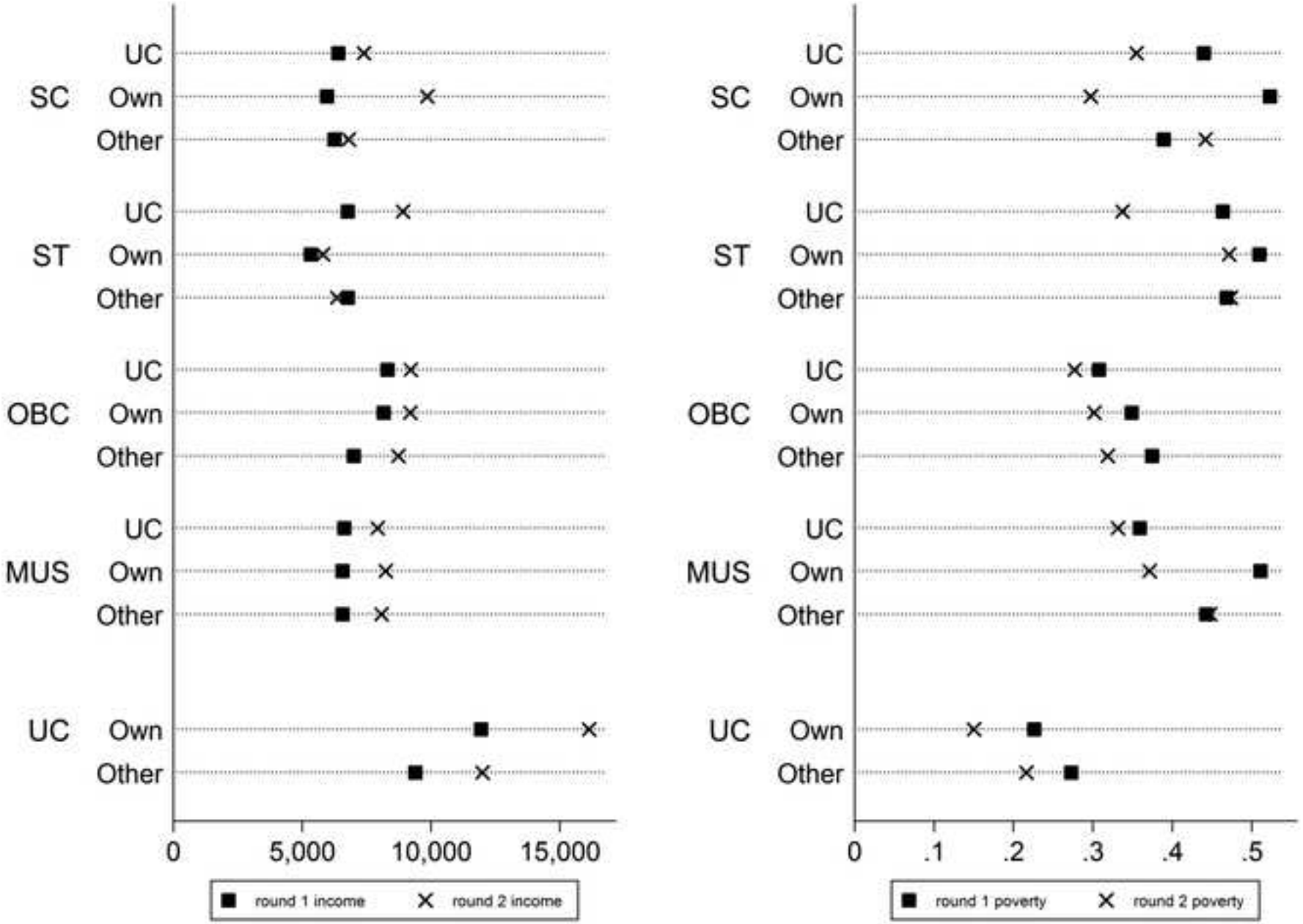

Figure
Click here to download high resolution image