WORKING PAPER NO: 618

Diversity Deficit and Scale-Flip

Naveen Bharathi
Harvard University
Cambridge, USA
naveen.bharathi@fac.harvard.edu

Deepak Malghan
Indian Institute of Management Bangalore
Bannerghatta Road, Bangalore – 5600 76
dmalghan@iimb.ac.in

Sumit Mishra
Krea University
Andhra Pradesh
sumit.mishra@krea.edu.in

Andaleeb Rahman
Cornell University
Ithaca NY 14853, USA
ar687@cornell.edu

Year of Publication – July 2020

Diversity Deficit and Scale-Flip☆

Naveen Bharathi[a], Deepak Malghan[b], Sumit Mishra[c], Andaleeb Rahman[d]

[a]*Harvard University*
[b]*Indian Institute of Management Bangalore*
[c]*Krea University*
[d]*Cornell University*

Abstract

We present a comprehensive multi-scale test of the diversity-deficit hypothesis that posits a negative association between diversity and development. We develop a"scale-flip hypothesis" that formalizes how the political salience of diversity is contingent on the level of analysis. We also contribute to the political economy of public goods literature using the largest dataset used to date — $n \approx 1.2$ million village-year points from a two-period panel of all villages in the Indian national census data. We find evidence for "scale-flip" so that there is a robust positive association between diversity and public goods at the local level.

Key words: Scale-flip Hypothesis, Public Goods, India, MAUP

☆We thank Adam Auerbach, Christian Davenport, Satish Deshpande, James Fearon, Devesh Kapur, Alexander Lee, Vijayendra Rao, Hillel Soifer, and Emmanuel Teitelbaum for useful comments and suggestions. The usual disclaimers apply. Bharathi acknowledges support from Harvard University's South Asia Institute. Part of the research reported here was carried out when Malghan was affiliated with Princeton University. He thanks the Princeton Institute for International and Regional Studies for funding support. Mishra received funding for this research from the International Center of Advanced Studies: Metamorphoses of the Political (ICAS:MP), in its TM6 (political economy) module.

Email addresses: naveen.bharathi@fas.harvard.edu (Naveen Bharathi), dmalghan@iimb.ac.in (Deepak Malghan), sumit.mishra@krea.edu.in (Sumit Mishra), ar687@cornell.edu (Andaleeb Rahman)

1. Introduction

The detrimental effect of social divisions has been termed as "one of the most powerful hypotheses" in the political economy of development (Banerjee et al., 2007). Led by seminal works of Easterly and Levine (1997) and Alesina et al. (1999), an impressive array of scholarship has found disparate empirical evidence for this diversity deficit — especially, the negative association between diversity and public goods provision (Alesina et al., 1999; Miguel and Gugerty, 2005; Banerjee and Somanathan, 2007). The diversity deficit thesis has also found empirical support across a range of outcomes beyond public goods provision including increased conflicts (Collier and Hoeffler, 1998; Collier, 2004; Montalvo and Reynal-Querol, 2005); reduced social trust (Alesina and La Ferrara, 2002; Dinesen and Sønderskov, 2015); social cohesion (van der Meer and Tolsma, 2014); and quality of governance (La Porta et al., 1999; Alesina and Zhuravskaya, 2011). These empirical findings are rooted in theories which emphasize inter-group cohesion (or lack thereof), preference heterogeneity, and intra-group social sanctioning, as the mediating channels between diversity and development (Habyarimana et al., 2007).

More recently however, comparative politics literature has mounted a theoretical and empirical challenge to the diversity deficit hypothesis by questioning the suitability of sovereign nation states as an appropriate level of analysis (Singh and vom Hau, 2016). Mixed evidence from empirical tests of the diversity deficit thesis at the sub-national level has added further ballast to these criticisms (Gerring et al., 2015; Gisselquist et al., 2016). Despite this accumulated empirical base, there is no agreement on whether a universal theory of inter-group relations can describe the effect of diversity across political and administrative levels.[1] The appropriate political level of analysis remains an open question.

We address the political levels of analysis problem by presenting, what is to the best of our knowledge, the first true large-scale multi-level test of the diversity deficit hypothesis. Unlike previous studies that combine multiple datasets and contexts (Gerring et al., 2015), we aggregate identity information for the *same* set of individuals into three different geographic, administrative, and political sub-national aggregations.[2] We use the two latest available Indian national census data that contains aggregate caste category information for over 800 million rural residents. Caste is the principal axis of stratification in rural India, and census-defined categories represent politically salient groups. We aggregate census data from $\approx$ 0.6 million villages, into $\approx$ 6,000 sub-districts that are contained within $\approx$ 600 districts. We use the four most significant local public goods (primary school, tap water, paved road, and electricity) at the village level as our dependent variables.

Our multi-scale analysis shows how the association between diversity and public goods is sensitive to political level of analysis. Diversity computed at aggregate levels can hide significant local variation. We show that districts — often the level of analysis in sub-national empirical exercises — may not be the most appropriate choice when public goods are provided in decentralized, or even a partially decentralized manner. The empirical findings are consistent with our argument that the study of group dynamics must be embedded in the political-geographical space that these social groups inhabit. In the Indian context, we challenge the current scholarship that posits a negative association between caste diversity and public goods. We show why not accounting for the most appropriate political level of analysis can result in a mischaracterization of the relationship between diversity and public goods. Our findings provide robust evidence for diversity dividend for public goods provisioning at the local level.

2. Spatial Aggregation and the Scale-Flip Hypothesis

The diversity deficit hypothesis emerged from cross-country regressions that suggested a negative relationship between diversity and economic performance (Easterly and Levine, 1997). Cross-country regressions have been criticized for neglecting potential endogeneity concerns rooted in the fact that both current development outcomes as well as ethno-linguistic diversity can be explained by common macro-historical processes of nation state formation (Wimmer, 2015). Empirical results from the subsequent 'sub-nation turn,' however, have been inconclusive or have even reported a positive association between diversity and development. Gisselquist et al. (2016) find public goods to be positively associated with diversity at sub-national levels in Zambia. Using cross-country micro-data on diversity and human development outcomes (at the level of a country, sub-national region, and district), Gerring et al. (2015) find this association to be sensitive to to political levels of analysis – diversity deficit at country level turns to diversity dividend at the district-level. Using a much finer geographic resolution where diversity is computed across various geospatial grid-sizes, Montalvo and Reynal-Querol (2017) show a positive association between diversity and growth and argue that potential for trade leads to greater cooperation at the local level.[3]

2.1. Scale-Flip Hypothesis

The deficit hypothesis is predicated on a variety of coordination failures in diverse societies. However, there are no theoretical reasons to assume that such coordination failures are independent of specific scale-contingent socio-political and institutional configurations. Contact theory suggests that diversity ceases to be a hindrance for cooperation when sustained inter-group interactions lead to increased trust and coordination (Allport et al., 1954; Pettigrew, 1998). In contrast, conflict theory posits an increase in tension when groups compete for limited resources (Blumer, 1958; Blalock, 1967). While these theories apparently contradict each other, "contact" and "conflict" outcomes are in practice actualized at different geographic scales (Bowyer, 2009; Kasara, 2013). The predictions of contact theory require regular interactions across group boundaries. Such contact is more likely at the local scale (for example, village or neighborhood).[4]

If coordination failures are scale-contingent, are there institutional configurations that can mitigate diversity deficit, and perhaps even *flip* diversity deficit into diversity dividend? Multitude of institutions working in consonance — for example, democracy, independent judiciary, civil society groups, free press, and consociational channels of communication — are essential to integrate different groups and provide them voice in order to harness their collective skills and coordinate differential preferences, at the national or sub-national level (Gerring et al., 2015). It is, therefore, fair to hypothesize a "scale-flip" so that these institutional Goldilocks conditions emerge at some sub-national scales but not at others. With a scale-flip, the political levels of analysis problem assumes even greater salience.

The appropriate political level for studying the effect of sub-national diversity is also mediated by the degree of constitutional authority and accountability which rests with the local government. If the local government has little control over resources, then it is possibly not a good level to study the political effects of diversity. Countries such as Bolivia, Brazil, Indonesia, and South Africa are examples where decentralization has been accompanied with greater legislative authority (Bardhan and Mookherjee, 2006). In contrast, Uganda and China devolved administrative, but not political authority to the local governments. While India and Pakistan have formally devolved political authority through local elections, the degree of state and elite influences has ensured that decentralization remains incomplete at best. A diverse literature on sub-national politics in India (the empirical context of this paper) has shown why the 'level of analysis' problem is politically salient (Chhibber and Nooruddin, 2004). Expenditure on public goods and the nature of redistributive politics are salient at specific sub-national scales (Saez and Sinha, 2010; Thachil and Teitelbaum, 2015). The politically relevant scale for studying public goods politics is also contingent on the characteristics of the specific public good in question. For example, national highways or irrigation systems (with high spillover effects) are generally the prerogative of the central or federal governments, while local roads, or canals are more likely to be determined by the strength of local coalitions (Mookherjee, 2015).[5]

Formal democratic institutions at local scales also engender deliberative platforms that diverse groups are able to use to voice their concerns and solve potential coordination problems (Sanyal and Rao, 2018). Common pool resources are the oft-cited example of communities transcending social and economic differences (Ostrom, 1990; Thapliyal et al., 2019). Sharing common spaces despite obvious markers of differences in culture or status fosters peaceful coexistence when there is sufficient civic engagement that softens group divisions. At larger geographic aggregations, however, there is competition among groups because in-group affinity or homophily is invariant to dwelling space. In terms of public goods, individuals at larger spatial aggregates remain unaware of actions by out-group individuals. When group members remain anonymous, inter-group aversion underpinning diversity deficit is salient (Belmonte et al., 2018).

2.2. Intra-unit Segregation

From an empirical perspective, the political level of analysis is crucial because country-level group configurations – or even those at higher-order sub-national units – are different from local group shares. Groups are often isolated and spatially segregated locally (for example, neighborhoods, villages, wards, census tracts). Spatial segregation impacts group cohesion (Robinson, 2017), and provides conditions for in-group bias in public goods placement (Ejdemyr et al., 2017). Regional concentration of specific groups also lead to localized pockets of highly polarized conflict-prone regions which a diversity metric at an aggregate level fails to capture (Bleaney and Dimico, 2017). This leads to well-known statistical inference issues when spatially aggregated data is used to test local hypotheses. (Robinson, 1950; Openshaw, 1984; King, 2013). For example, diverse cities often contain segregated neighborhoods, like in the United States where neighborhoods are racially segregated.[6] Therefore, when the level of analysis is neglected, it is possible to obtain "almost any desired result by aggregating the data in different ways" (Fotheringham and Wong, 1991).

2.3. Politics of Scale-Flip

The theoretical and empirical import of the levels of analysis problem is ultimately rooted in the fact that politics itself is scale-contingent. Contrary to early theoretical explanations for diversity deficit that assumed a bottom-up creation of public goods, top-down state provisioning better explains spatial inequality in public goods at the local scale (Banerjee et al., 2007). Public good politics must describe when (if) communities are able to effectively overcome collective action problems, *and* the state's response to community petition for public goods (Tajima et al., 2018). Petitions for public goods are scale-contingent, so that when empirical analysis aggregates distribution of public goods, politically salient local variations are overlooked. Even more crucially, aggregate analysis cannot describe the politics of citizen-state relationship in placement of public goods. In the top-down provision of public goods, devolution of funds goes through multiple layers of administration, each with sufficient residual discretion to indulge in discrimination or favoritism in spatial placement of the good (Besley et al., 2004; Ejdemyr et al., 2017; Lee, 2018).

Many developing countries including India have a decentralized administrative structure, where public goods are provided by the state through multiple layers of a nested administrative network that includes bureaucrats, local actors and civil society. The partially decentralized state apparatus with sufficient discretion often indulges in spatial discrimination (Bardhan, 2002). When the state is able to spatially discriminate, politics at the local level – where the goods are actually placed – is at least as important as politics at higher levels. Statistical problems that plague aggregation of local diversity at higher levels also impact measures of public goods (e.g., literacy rates or social sector expenditure) by masking their spatial spread and the underlying political process.

Even when the emerging sub-national literature has focused on local public goods, it has not accounted for how empirical results are potentially contingent on political levels of analysis. We argue that in the context of rural India, it is indeed fair to hypothesize a scale-flip — an apparent diversity deficit at large sub-national aggregates such as districts flip to a diversity credit so that local diversity is actually positively associated with local public goods. While the scale-flip hypothesis is perhaps not necessarily generalizable, any evidence for scale-flip does underscore the scale-contingency of the association between diversity and public goods.[7]

3. Diversity and Public Goods Politics in Rural India

The institutional landscape of rural India is a fecund empirical site to investigate the scale-flip hypothesis. Caste is widely recognized as the most important social cleavage, as well as the principal axis of stratification in rural India. Status-seeking behavior among hierarchical caste groups is common and caste groups can develop "spiteful preferences" towards each other that inhibits inter-group cooperation and impacts overall welfare (Fehr et al., 2008). Elected representatives often exhibit a "taste for discrimination" against out-group constituents (Becker, 1957). Conflicting preferences between caste groups are manifested spatially. For example, groups often disagree over the location of public goods such as the location of a public tap or a well (Munshi and Rosenzweig, 2016).

Given the decentralized nature of Indian polity, local elected representatives resort to preferential treatment towards their own community, or village when it comes to spatial placement of public goods (Besley et al., 2004). Elementary endogamous groups (*jatis*) are important for characterizing politics in rural India. However, a particularly important group boundary for spatial discrimination is the demarcation between formerly so-called "untouchable" caste groups (officially classified as SCs, or "Scheduled Castes") and others.[8] India's indigenous tribes (administratively classified as "Scheduled Tribes," or STs) are at least as marginalized as the SCs but their demographic spread gives them little political voice. Our analysis therefore uses the politically salient boundaries between *Dalits*, *Tribals*, and other social groups.

3.1. A Case for Distinguishing Levels of Analysis

Banerjee and Somanathan (2007) inaugurated the empirical analysis of Indian public goods by reporting a negative association between caste diversity and an array of public goods in rural India using $\approx$ 500 large parliamentary constituencies as their level of analysis. They find that greater population share of *Brahmins* (group at the apex of the traditional caste hierarchy) is however positively associated with public goods such as school, post office, and piped water.[9] Instead of assuming that a particular level of political analysis (say, a parliamentary constituency) is the most appropriate, we make the case for a theoretically and empirically grounded choice. Data limitations notwithstanding, the political unit of analysis should be informed by what the salient political processes are at any given level. The use of large political aggregates such as parliamentary constituencies, ignores the fact that it can contain many homogeneous and segregated villages, leading to biased inferences. Such bias introduced by segregation is even more pronounced when one accounts for the stratification produced by the hierarchical caste order (Lee, 2018).

The case for not ignoring the village as an empirical level of analysis is also motivated by contextual institutional and political configurations. Public goods in rural India are allocated in a decentralized manner through village councils, known as *Gram Panchayats* (GPs). GPs have considerable discretion in the placement of public goods, such as piped water, school or health centers. GPs, which represent a cluster of villages, form the lowest tier of the elected government structure and are responsible for essential infrastructure, identifying beneficiaries of welfare programs, and even adjudicating local disputes.

The 73rd amendment to the Indian constitution in 1993 ushered in an *elected* three-tier local government structure across the country (Singh, 1994). This nested tiered rural administrative structure is embedded within the Indian federal structure, and the multiple levels of the state interact to provide public goods (Chhibber et al., 2004). At the lowest level, funds are funneled to the villages through GPs, sub-districts and other higher tiers facilitated by the interactions between bureaucrats and politicians across these administrative tiers of governance. Politics in these "local democracies" is dominated by the forces of patronage, or identity based favoritism, which influences public goods placement (Heller et al., 2007; Bardhan et al., 2009). The partially decentralized governance structure provides incentives for elected leaders to indulge in in-group favoritism or out-group discrimination — and these incentives are more acute when villages are segregated. Dominant groups can use their political power to undermine public good provisioning in villages inhabited by the marginalized groups (Lee, 2018).

3.2. Local Group Dynamics and Potential Scale-Flip

Local public goods are allocated in India within a nested administrative hierarchy, with varying degrees of decentralization (Banerjee et al., 2007; Bardhan, 2002). When the state provides public goods and local politicians are elected, group diversity implies more fragmented social networks which reduces the likelihood of elite capture (Cruz et al., 2020). Diversity, therefore, can have a positive association with public goods. Decentralized administrative system brings the state closer to the citizens so that policy outcomes are more representative and not captured by elite preferences (Bandiera and Levy, 2011). Greater coordination across caste groups (needed to effectively lobby the state for public goods) is further enhanced when quotas are mandated in village council seats for the marginalized caste groups. Greater political participation and shared local leadership lead to more frequent inter-group interactions that can ameliorate discriminatory practices in a stratified caste structure (Chauchard, 2014).[10]

Skill complementarity, resulting in greater trade possibilities, in India, has been found to foster greater inter-group coordination (Jha, 2013). Caste (with a significant overlap with class) is linked to traditional hereditary occupation. At the village level, the relationship between caste groups is reciprocal even if unequal. While politicization of caste boundaries does lead to divisive behavior, it does not represents a state of constant conflict. Even in the absence of dense social networks that span group boundaries, complementary occupational skills provide a pathway for generation of public goods — not unlike metropolitan centers in advanced economies that are segregated by socio-economic or racial markers, but also represent complementary means of economic production. Even while villages are not as heterogeneous as cities, the skill diversity channel is important when forces of democracy embolden marginalized groups to raise voice and enjoin other groups to demand public goods, leading to diversity dividend at the local level.

Finally, among the homogeneous villages inhabited by marginalized groups (SC/ST), public goods are expected to be lower on account of their historical discrimination by the political elites, despite reservation policies (Lee, 2018). For the homogeneous OTH villages, on the other hand, demand for public goods is undermined by possible factional divisions among the larger caste, which has a similar effect on undermining their social capital to lobby for the public goods (we control for both these channels in our empirical exercise).

Local-level democracies, the village assemblies also serve as crucial sites of deliberation and consultation, potentially ameliorating the diversity penalty related to coordination problems (Sanyal and Rao, 2018). Village assemblies (*gram sabhas*) in India meet regularly, and even with incomplete decentralization, these meetings draw up lobbying or petitioning strategies for collective welfare, especially public goods. Studying caste-fragmented villages in South India, Ban et al. (2012) show that these local village democracies are not merely 'talking shops,' but are sites for voicing welfare concerns to a credible agent – the local elected representative. Formal local democracies complement informal caste-based local institutions in Indian villages (Ananth Pur, 2007). The informal local institutions, over time, have adapted to the competitive political environment generated by the formal village councils, and exhibit greater inclusiveness (Ananth Pur and Moore, 2010). Democratic representation of the marginalized caste groups has also helped in mobilizing support for public goods and services. When local politicians possess discretionary powers to determine public goods placement, they also pursue opportunities for private profits, which includes electoral gains or continuity in office. Local elites, therefore, have an incentive to distribute public goods equitably — mitigating any baleful effects of diversity (Bohlken, 2016).

4. Data

In order to perform a true multi-scale test of the diversity deficit, and scale-flip hypotheses, we use the decennial Indian national population census data for years 2001 and 2011 (two latest census years). We use the `Primary Census Abstract (PCA)` for demographic data; and amenities information from `Series H`, and `Village Directory (VD)` tables for public goods. We combine these datasets for each census year and geographical aggregation – village, sub-district, and district (Table 1). We use incidence data for four of the most basic public amenities – primary school, tap-water facility, paved road, and electricity provision – as our public goods measures. From the binary indicator variable at the village level, we aggregate them into share of villages with public goods at the sub-district and district level — a standard practice in extant empirical literature (Banerjee and Somanathan, 2007; Banerjee et al., 2005). Table 2 summarizes village-level public goods.

The main explanatory variable – census category diversity – and other covariates come from the `PCA` which also contains additional information on occupational structure, gender distribution, population, and literacy status of the population, which we use as control variables. In absence of detailed caste-structure information in the census data, we rely on the aggregate census caste categories: Scheduled Caste (SC), Scheduled Tribe (ST), and the residual all other castes grouped as Other Castes (OTH). SCs and STs are the most marginalized social groups in India. It must be noted that the Indian caste structure is far more complex than these aggregate census categories. However, census classifications not only represent the most significant caste boundaries – demarcating the historically "untouchable" and indigenous groups from the rest of the population – but also form the basis of most administrative policies including affirmative action and targeted welfare policies (Deliege, 1992; Guru, 2009). Affirmative action policies in favor of the SCs and the STs have entrenched these meta-classifications politically and have hastened the convergence between administratively constructed group boundaries and politically salient cleavages (Jaffrelot, 2006; Htun, 2004).[11]

Indian caste system is built on the foundation of endogamous ascriptive caste markers (*jatis*) which collapse into these politically salient major categories despite obvious differences in hierarchical status (Gupta, 2005). However, the national census does not collect *jati* data. A common practice in the empirical literature has been to use *jati* data from the 1931 colonial census (the last time such data was tabulated) as an instrument for current level of caste diversity. We refrain from using the extrapolated 1931 census data as a proxy for contemporary diversity because caste data collected in 1931 is likely an unreliable measure of present-day diversity – several decades later that when the intervening period includes the massive displacement of people during the partition of the subcontinent. The use of 1931 caste census data requires a number of other additional assumptions. For instance, Banerjee and Somanathan (2007) assume that the fertility rates, and migration rates remain the same across caste-groups over time. These untestable assumptions can potentially bias econometric estimations (Kelly, 2019).

Even more centrally from the perspective of this paper, disaggregated *jati* data for 1931 is available only at the district level.[12] No *jati* data is available at the village level in the 1931 census tables. The primary purpose of our empirical exercise is to investigate the political level of analysis problem, and determine how (if) it must be accounted for in tests for diversity deficit hypothesis, *given* a specific politically salient group boundary. While a complete characterization of the relationship between caste diversity and public goods will need village-level *jati* data, our principal goal is well-served by accounting for the most significant boundaries represented by census categories. The central argument of this paper is primarily contingent on high-resolution spatial data which are not theoretically invalidated by the lack of high-resolution caste category information.

5. Diversity Cascade

We construct the workhorse diversity metric — fractionalization index (ELF) at three cascading levels of spatial aggregation — $z \in \{\text{village, sub-district, district}\}$.

$$FRA_z = 1 - \left(\sum_{\forall\, c \in z} \pi_c^2 \right) \tag{1}$$

where π_c is the population share of caste groups ($c \in \{SC, ST, OTH\}$), in the geographical region z. FRA_z represents the probability that if two individuals are chosen randomly from the population, they would be from distinct social groups.

There is sufficient variation in SC/ST population at all three levels of analysis (village, sub-district, and district) that justifies the use of $c \in \{SC, ST, OTH\}$, rather than collapsing SC and ST into a composite SCST category representing historically marginalized groups. This variation is shown in Figure 1.

5.1. Diversity Across Geographic Scales

The largest sub-national unit in India's federal polity – the states – are large aggregations containing several districts, sub-districts, villages, and towns. In 2011, 18 states had a population of over 25 million, and the largest state (Uttar Pradesh)

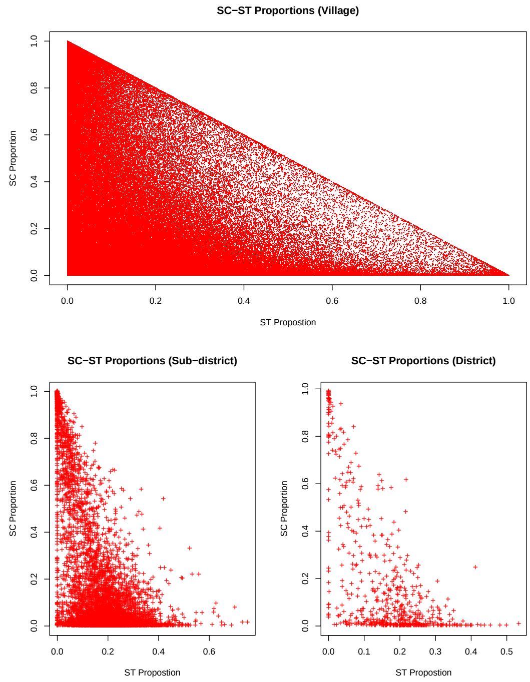

Figure 1. SC–ST Population Overlap

had a population of nearly 200 million. As a point of comparison, in 2011, 10 Indian states had populations exceeding that of France. While Indian states are linguistically homogeneous they contain diverse social groups that makes sub-national politics in India complex (Chhibber and Petrocik, 1989; Laitin, 1989). Indian states (Singh, 2015b,a), and districts or parliamentary constituencies that largely overlap districts (Banerjee and Somanathan, 2007; Balasubramaniam et al., 2014) have been the sub-national levels of analysis that have dominated the literature.

5.2. Spatial Aggregation of Diversity

Figure 2 shows how potential ecological fallacy problems plague diversity metrics computed at large aggregate sub-national units in India. For the twenty-five largest states in India, we plot the distribution of FRA_z, at district, *taluka* (sub-district) and village levels.[13] The figure also shows mean village and district fractionalization index for each state (the dashed lines show village means and solid lines represent district means). The obvious differences between village and district distributions in Figure 2 represent more than a statistical artifact related to number of observations (the village distribution comes from many more observations than district or sub-district distributions). Figure 2 illustrates how potential statistical inference problems in computing diversity metrics at large aggregations such as districts or sub-districts can bias empirical estimates. The aggregate diversity metric that describes a diverse district can hide the fact that individual villages might themselves be homogeneous and spatially segregated. If villages are the actual physical sites where public goods are placed, Figure 2 shows that testing the diversity deficit hypothesis at an aggregate level (such as districts or sub-districts) can potentially bias empirical models.

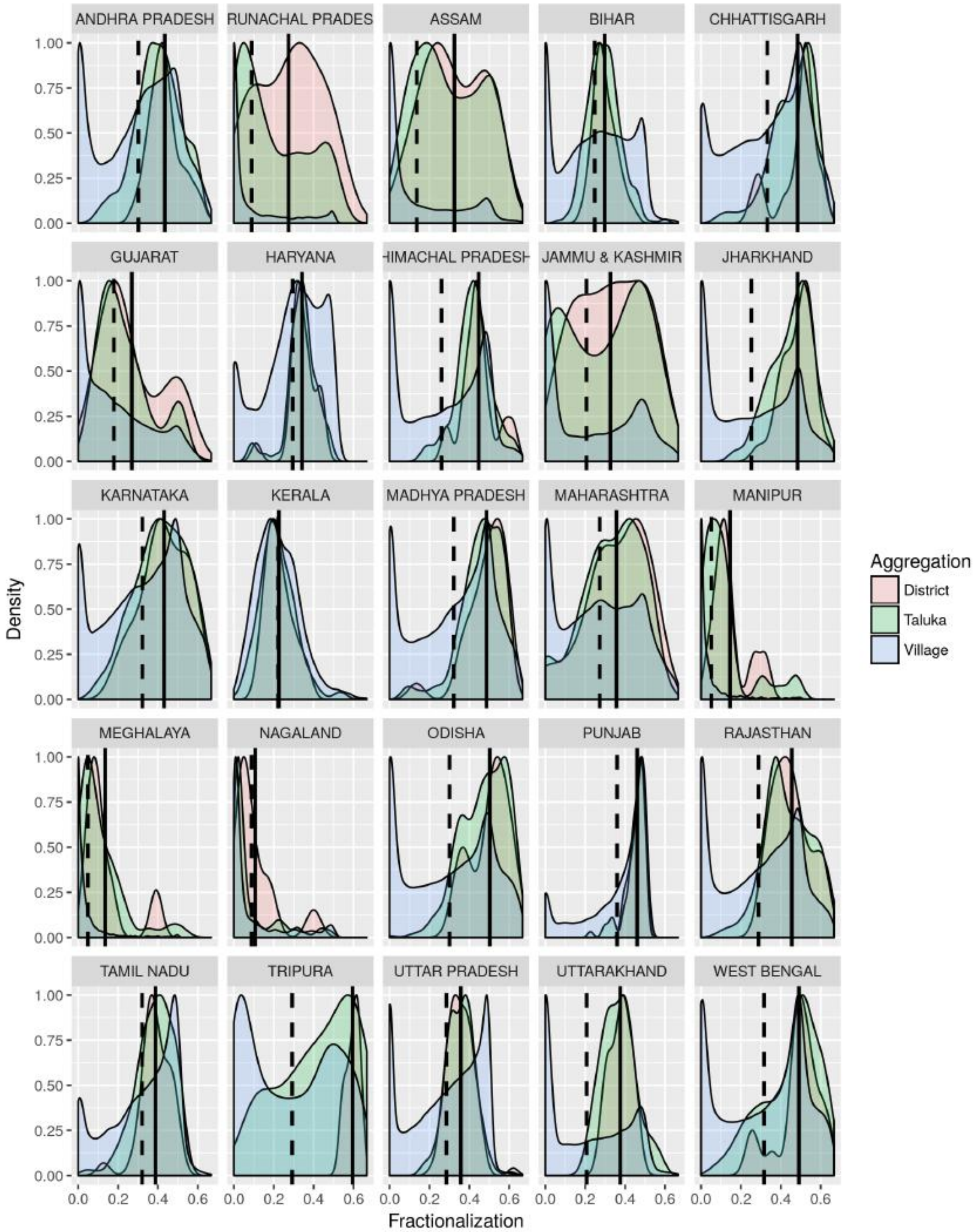

Figure 2. State-wise distribution plots for fractionalization by geographic aggregations. See main text for explanation.

Figure 2 also shows that there is considerable variation in the severity of potential statistical inference problems related to the use of aggregate diversity metrics. In highly populated states like Uttar Pradesh (UP) or Bihar, the village level distribution is clearly bimodal (even as district and village averages are closely bunched). On the other hand for the south Indian state of Kerala, the fractionalization metric is normally distributed across all three geographic aggregations and the three distributions are nearly identical. Thus, sub-national comparisons of human development and welfare outcomes across different states of India (Singh, 2015b,a) are potentially biased by how ecological inference problem differentially impacts the states that are being compared.

5.3. Aggregating Local-level Public Goods

Biases related to the statistical inference problem are not only introduced by how diversity is measured at aggregate levels, but also by how public goods themselves are distributed across villages. In the empirical literature on India, the public goods variable is defined as the percentage of villages in a district, parliamentary constituency, or some other higher aggregation. When the actual physical location of the public good is a village and the empirical level of analysis is some higher aggregation, ecological inference problems are introduced when there is sufficient intra-unit spatial variation in public goods incidence. Figure 3 illustrates this problem. The figure shows variation of public goods incidence within a single randomly chosen district (Allahabad) form the largest state in India, Uttar Pradesh. The figure shows village as well as sub-district boundaries. There is not only spatial variation within the district but even within a sub-district, across all four public goods (school, road, water, and electricity). Taken together, aggregation biases in the measurement of both public goods as well as diversity calls into question previous studies of the association between diversity and public goods in India.

6. Empirical Model, and Results

The principal multi-scale specification that we use to test the diversity deficit and the scale-flip hypotheses is:

$$Y_i^z = \alpha \; + \; \beta \cdot FRA_i^z \; + \; \gamma \cdot \mathbf{X}_i^z \; + \; \lambda \cdot \mathbf{FE}_j^{Z \succ z} \; + \; \epsilon_i^{Z \succ z} \tag{2}$$

where Y_i^z is the measure of public good Y in the ith unit at a given level of aggregation, z. We run the model at three different aggregations – $z \in$ {Village, Sub-district, District}. At the village-level Y is a binary incidence variable, and at the sub-district and district levels, Y represents share of villages with the public good. FRA_i^z is the standard fractionalization index (Equation 1), computed at the appropriate aggregation level. $\mathbf{X}_i^z$ is a vector of control variables at aggregation z. Our baseline models control for total population, literacy rate, sex ratio, and workforce participation rates at all three aggregations. Additionally, the sub-district and district level models also control number of villages. As a base robustness-check, we use data from both 2001 and 2011 national census to estimate the model in (2) at each of the three aggregation levels.

Models at each of the three aggregations include the immediate higher aggregation ($Z \succ z$) as a fixed effect. Thus, $\mathbf{FE}_j^{Z \succ z}$ corresponds to the sub-district j containing the village i as a fixed effect in village-level regressions; district containing the sub-district in sub-district level regressions, and the state containing the district in district level models. These fixed effect terms not only control for segregation, but a host of other variations like agro-ecological zones, agricultural productivity and regional variations in economic development. The use of sub-district fixed effect also accounts for colonial

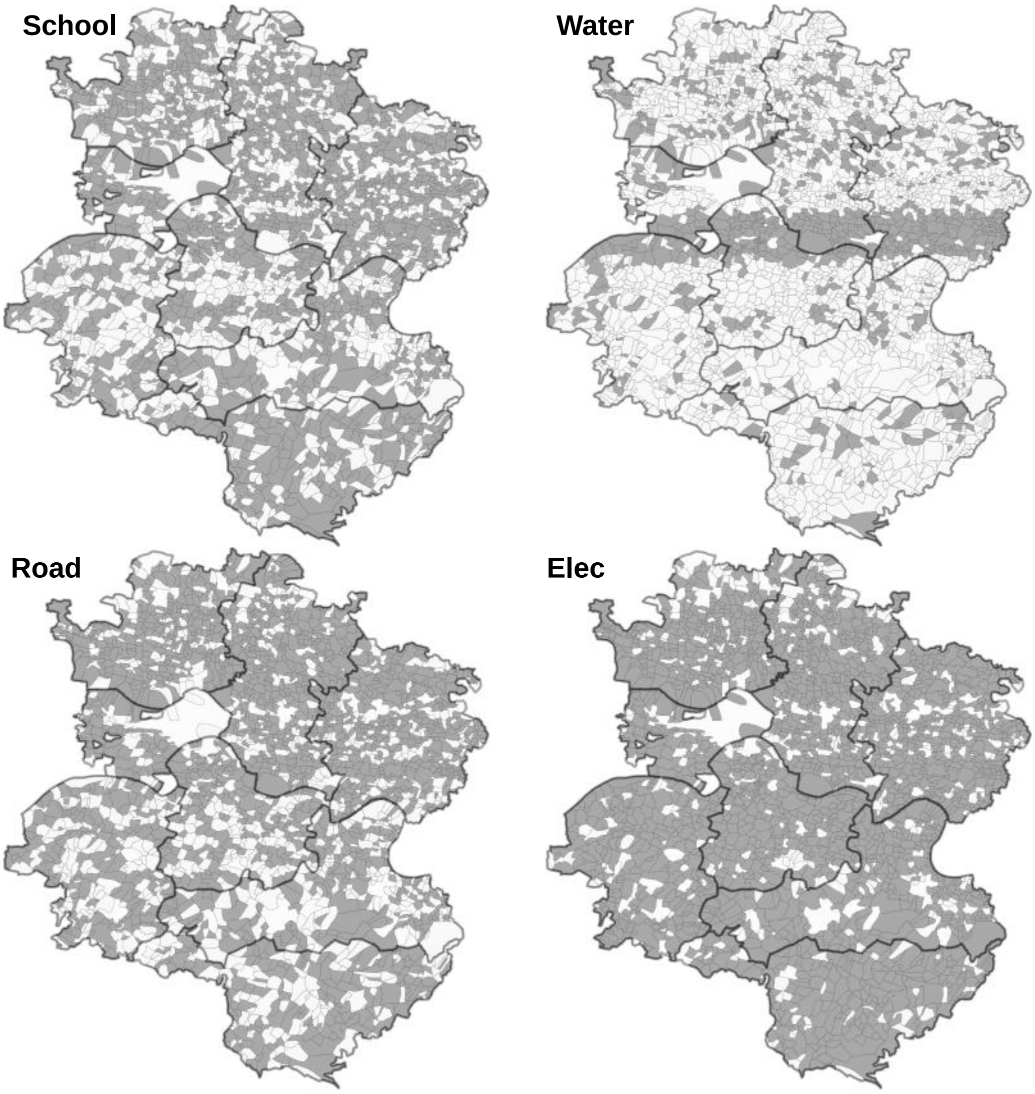

Figure 3. Public good across villages, within the subdistricts of *Allahabad*, a district in the state of Uttar Pradesh

legacy and its role in economic development and public good provisioning (Banerjee et al., 2005; Banerjee and Iyer, 2005). Each of our regression models also control for historical group hierarchy, and spatial segregation, through including group population shares of the marginalized groups, SCs and STs. The standard errors ($\epsilon_i^{Z \succ z}$), are clustered at the appropriate successor scale, ($Z \succ z$).

6.1. Results

Coefficients on the FRA^z from the suite of models at different spatial aggregations are reported in Table 3. Consistent with our argument, the effects of diversity on public goods is not robust across the three political levels of analysis, regardless of the choice of public good. Indeed, we find evidence for the scale-flip hypothesis. At the village level, we find that diversity is positively associated with all four public goods. However, at the sub-district level, diversity has a statistically significant relationship with only two of the goods. Share of villages with a primary school decrease with sub-district level diversity, while share of villages with electricity provision shows the reverse association. At district-level, we find none of the public goods (except tap-water) have a statistically significant association with diversity.

Additionally, we also estimate the effect of diversity – at each level – on the change in the public good between 2001 and 2011, and once again find evidence for scale-flip. Diversity is positively associated with public goods at the village-level (Table 4). However, sub-district diversity has a negative relationship with change in the share of villages with a primary school or tapwater facility. The result is reversed when the share of villages with electricity is the outcome variable of interest. The association between district-level diversity and public goods is not statistically significant.

Taken together, results in Table 3 and Table 4 provide strong support for the scale-flip hypothesis — with the association between diversity and public goods moving from being positive, to negative, to being statistically insignificant at the highest aggregation. Our results challenge existing diversity deficit findings in the Indian context.

6.2. Migration and Optimal Sorting Concerns

A natural assertion in the scholarship on diversity-development association has been the concerns of endogeneity; individuals choose to migrate into more developed regions which affects regional diversity levels. Empirical models of diversity and development – especially in a sub-national setting – can suffer from potential endogeneity concerns arising from the fact that individuals can choose to migrate to more developed regions. Such 'optimal sorting' concerns are further compounded with divergent spatial mobility patterns across subgroups. However, optimal sorting is likely not a concern in the rural Indian context. The correlation coefficient between the fractionalization index over the period 1991-2001-2011 at all the three levels of aggregation – village, sub-district and district – are at least 0.9.[14]

The temporal stability of rural diversity in India is consistent with findings elsewhere.[15] Well-established findings from the migration literature further support our results. India has one of the lowest rates of permanent migration across the developing world. Workers who migrate only during the lean agricultural season and are counted as residents of their villages (Keshri and Bhagat, 2012). The most significant contributor to migration in India is marriage related mobility, among women (Rosenzweig and Stark, 1989). Given that marriage within a caste group is the norm, it is unlikely to bias our results. State boundaries further act as a hindrance to migration — social security benefits do not travel well across state boundaries, and India's linguistic states that each use mutually unintelligible official language acts as an additional barrier to inter-state mobility (Kone et al., 2018).

6.3. Additional Robustness Checks

The supplementary materials contain full-model results (including coefficients on our primary control variables) as well as several robustness checks.

Loss of Public Goods

We extend the analysis around change in public goods discussed above (Table-4) to include cases where a village has lost access to a public good. There is no theoretical justification for leaving out such villages — if diversity is associated with public goods change, we must also test for the possibility that it is implicated in loss of public goods. We present two sets of models around public goods change in supplementary materials that both confirm the robustness of our central results. First, we explicitly model gain and loss of public goods. Second, we separately analysed those villages which lost access to a public good. Our results suggest that more diverse villages were less likely to lose access to primary school, road, or electricity — once again supporting the arguments made in the paper.

Other Backward Castes

The empirical results presented here are based on Indian national census data that reports caste information as three broad aggregates – SC, ST, and a residual 'Others.' As we have argued in the paper, these (administratively) constructed categories are politically salient. However, the residual "others" category that accounts for 75.6% of the population is very diverse and includes a wide spectrum of caste groups from dominant "upper caste" groups to marginalized peasant groups (administratively classified as Other Backward Castes, OBCs). We used the 68th Round of the National Sample Survey (NSS) data to obtain district OBC-shares. After matching census districts with weighted NSS proportions, we ran sub-sample regressions for districts with OBC proportions in the top and bottom quartiles at all three geographic aggregations and find that our central findings are robust.[16]

Temporal Stability

As we use data from two national census cycles (2001, and 2011), we want to be sure that our results are not being driven by change in group compositions. For example, does the differential rural-urban migration rates between subgroups impact our results? In the supplementary materials, we present correlation between SC/ST proportions (at all three levels of analysis) in 2001 and 2011; and also 2001-11 correlations between fractionalization indices at each of the three levels of geographic aggregation. These correlation tables make clear that our results are not driven by change in diversity between 2001 and 2011.[17]

Marginalized Groups

Not all equally diverse aggregations (village, sub-district, or district) are created equal. It is conceivable that a homogeneous village (or a sub-district, district) that is dominated by upper caste groups will have public goods outcomes that are different from equally homogeneous aggregations numerically dominated by marginalized groups (SC or ST). All the regression models presented here use SC/ST population shares as controls. We extend these models in supplementary materials to explicitly model the interaction between fractionalization and SC/ST proportions. Additionally, we also run sub-sample regressions (at each of the three aggregation levels) for two sub-samples corresponding to an aggregation dominated by SC/ST groups or others respectively. These results once again confirm the central findings reported here.[18]

7. Conclusion

Through an empirical illustration from rural India, we have shown that the appropriate choice of political level of analysis in diversity-development debates must be grounded in sound theory. We introduce the scale-flip hypothesis and find that the effect of diversity on public good is sensitive to political level of analysis. By aggregating the same set of individuals at different levels (village, sub-district, and district), our analysis transcends a major limitation in extant literature (Gerring et al., 2015). We also show how instability of results across levels is driven by potential statistical inference problems, which has been a blind spot in the diversity-development debate. While we have established that the natural level of analysis for studying local public goods is the village in the Indian context, it could be a different level elsewhere, because "the political level of analysis problem" is embedded in specific institutional regimes.

Beyond contributing to the broader literature on diversity and public goods, this paper has called into question specific claims about public goods politics in India that use large political aggregations such as districts or parliamentary constituencies (Banerjee and Somanathan, 2007; Balasubramaniam et al., 2014).

Notes

[1]For instance, inter-group violence has been studied across countries (Fearon, 2003), within sub-national boundaries (Wilkinson, 2006), cities (Varshney, 2003), counties (Tolnay and Beck, 1995), or even grid-cells (Pierskalla and Hollenbach, 2013). Similarly, studies of the association between diversity and public good provision have used nation states (Baldwin and Huber, 2010), legislative assemblies (Banerjee et al., 2007), cities (Alesina et al., 1999), or villages (Munshi and Rosenzweig, 2016; Miguel and Gugerty, 2005).

[2]Gerring et al. (2015), however, do include country fixed effects. *Cf.* Gerring et al. (2017) for a multi-scale test of status of minorities as a function of geographic unit of analysis. The authors use a subset of global Demographic Health Survey (DHS) data from nine different countries where DHS identifies two sub-national aggregations – regions and districts. However, this cross-national dataset relies on definitions of regions and districts that can vary across countries. We are grateful to an anonymous reviewer for pointing us to this article as an example of a true multi-scale test in a cognate field.

[3]Scholars in the field of comparative politics have also made a strong case for *diversity-dividend* at the sub-national level, albeit through multiple channels (Soifer, 2016; Gao, 2016; McDonnell, 2016; Singh, 2015b). However, these sub-national studies also suffer from problems related to theoretical and empirical concerns of scale-dependent group behavior and spatial aggregation effects (Soifer, 2019).

[4]The pertinent point here is not whether contact is possible at larger aggregations such as the district, but that contact is at the very least *potentially* contingent on scale. While sustained contact is more likely at local scales, the relationship between contact and scale is ultimately an empirical question. We are grateful an anonymous reviewer for helping clarify this linkage between contact theory and scale.

[5]Spatial variation in public goods also occur on account of regional partisanship and the nature of urban concentration, as in Ecuador and Colombia (Soifer, 2016). Also *cf* Trounstine (2016).

[6]See `https://www.washingtonpost.com/graphics/2018/national/segregation-us-cities/` (accessed, March 30th 2020).

[7]Also *cf.* Gerring et al. (2015) who report a scale-flip in a diverse set of countries around the world.

[8]While administratively classified as SCs, or "Scheduled Castes," many of these groups self-identify themselves as *Dalits*, or the oppressed

[9]Revisiting the same question, albeit with updated district-level census information, Balasubramaniam et al. (2014) found that households in districts with higher caste diversity have lower access to tap water while religious diversity exhibits an opposite effect.

[10]Measuring the psychological effect of local reservations for the SCs and STs, Chauchard (2014) finds that while upper-caste individuals may continue to harbor biases against marginal groups that avail quotas, inter-group interactions are amiable and there is a reduction in discriminatory practices.

[11]The 73rd Amendment to the Indian constitution (1992) that extended political quotas to local government elections has further transformed the nature of redistributive local political preferences, thereby affecting the provision of public gods (Besley et al., 2004; Parthasarathy, 2017).

[12]Highly aggregate category data that combines caste and religion is available at the sub-district level, too.

[13]For meaningful representation of the distribution at all three levels, we dropped states with less that thirty sub-districts (administratively, sub-districts are most commonly referred to as *Talukas*, or *Thesils*). Across India, $\approx 600,000$ villages are embedded in ≈ 6000 sub-districts, that in turn are contained within ≈ 600 districts.

[14]We calculate similar coefficients across each of the Indian states and the correlations continue to be very high. Also *cf.* §D of supplementary materials.

[15]For example, *cf.* Gershman and Rivera (2018) for data from Africa.

[16]§C, supplementary materials

[17]§D, supplementary materials

[18]§E, supplementary materials

References

Alesina, A., Baqir, R., and Easterly, W. (1999). Public Goods and Ethnic Divisions. *The Quarterly Journal of Economics*, 114(4):1243–1284.

Alesina, A. and La Ferrara, E. (2002). Who Trusts Others? *Journal of Public Economics*, 85(2):207–234.

Alesina, A. and Zhuravskaya, E. (2011). Segregation and the Quality of Government in a Cross Section of Countries. *The American Economic Review*, 101(5):1872–1911.

Allport, G. W., Clark, K., and Pettigrew, T. (1954). *The Nature of Prejudice*. Addison-Wesley Reading, MA.

Ananth Pur, K. (2007). Rivalry or Synergy? Formal and Informal Local Governance in Rural India. *Development and Change*, 38(3):401–421.

Ananth Pur, K. and Moore, M. (2010). Ambiguous Institutions: Traditional Governance and Local Democracy in Rural South India. *Journal of Development Studies*, 46(4):603–623.

Balasubramaniam, D., Chatterjee, S., and Mustard, D. B. (2014). Got water? Social divisions and access to public goods in rural india. *Economica*, 81(321):140–160.

Baldwin, K. and Huber, J. D. (2010). Economic versus Cultural Differences: Forms of Ethnic Diversity and Public Goods Provision. *The American Political Science Review*, 104(4):644–662.

Ban, R., Jha, S., and Rao, V. (2012). Who has Voice in a Deliberative Democracy? Evidence from Transcripts of Village Parliaments in South India. *Journal of Development Economics*, 99(2):428–438.

Bandiera, O. and Levy, G. (2011). Diversity and the power of the elites in democratic societies: Evidence from Indonesia. *Journal of Public Economics*, 95(11-12):1322–1330.

Banerjee, A. and Iyer, L. (2005). History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India. *American Economic Review*, 95(4):1190–1213.

Banerjee, A., Iyer, L., and Somanathan, R. (2005). History, Social divisions, and Public Goods in Rural India. *Journal of the European Economic Association*, 3(2-3):639–647.

Banerjee, A., Iyer, L., and Somanathan, R. (2007). Public Action for Public Goods. In *Handbook of Development Economics*, volume 4, pages 3117–3154.

Banerjee, A. and Somanathan, R. (2007). The Political Economy of Public Goods: Some Evidence from India. *Journal of Development Economics*, 82(2):287–314.

Bardhan, P. (2002). Decentralization of Governance and Development. *Journal of Economic Perspectives*, 16(4):185–205.

Bardhan, P., Mitra, S., Mookherjee, D., Sarkar, A., et al. (2009). Local Democracy and Clientelism: Implications for Political Stability in Rural West Bengal. *Economic and Political Weekly*, 44(9):46–58.

Bardhan, P. and Mookherjee, D. (2006). *Decentralization and Local Governance in Developing Countries: A Comparative Perspective*, volume 1. The MIT Press.

Becker, G. (1957). *The Economics of Discrimination*. Chicago Press, Chicago.

Belmonte, A., Dell'Anno, R., and Teobaldelli, D. (2018). Tax morale, aversion to ethnic diversity, and Decentralization. *European Journal of Political Economy*, 55:204–223.

Besley, T., Pande, R., Rahman, L., and Rao, V. (2004). The Politics of Public Good Provision: Evidence from Indian local governments. *Journal of the European Economic Association*, 2(2-3):416–426.

Blalock, H. M. (1967). *Toward a Theory of Minority-group Relations*, volume 325.

New York: Wiley.
Bleaney, M. and Dimico, A. (2017). Ethnic Diversity and Conflict. *Journal of Institutional Economics*, 13(02):357–378.
Blumer, H. (1958). Race prejudice as a sense of group position. *Pacific Sociological Review*, 1(1):3–7.
Bohlken, A. T. (2016). *Democratization from Above: The Logic of Local Democracy in the Developing World*. Cambridge University Press, Cambridge.
Bowyer, B. T. (2009). The Contextual Determinants of Whites Racial Attitudes in England. *British Journal of Political Science*, 39(3):559–586.
Chauchard, S. (2014). Can Descriptive Representation Change Beliefs about a Stigmatized Group? Evidence from Rural India. *American Political Science Review*, 108(02):403–422.
Chhibber, P. and Nooruddin, I. (2004). Do Party Systems Count? The Number of Parties and Government Performance in the Indian States. *Comparative Political Studies*, 37(2):152–187.
Chhibber, P., Shastri, S., and Sisson, R. (2004). Federal Arrangements and the Provision of Public Goods in India. *Asian Survey*, 44(3):339–352.
Chhibber, P. K. and Petrocik, J. R. (1989). The puzzle of Indian Politics: Social Cleavages and the Indian Party System. *British Journal of Political Science*, 19(2):191–210.
Collier, P. (2004). Greed and Grievance in Civil War. *Oxford Economic Papers*, 56(4):563–595.
Collier, P. and Hoeffler, A. (1998). On Economic Causes of Civil War. *Oxford Economic Papers*, 50(4):563–573.
Cruz, C., Labonne, J., and Querubin, P. (2020). Social network structures and the politics of public goods provision: Evidence from the philippines. *American Political Science Review*, 114(2):486–501.
Deliege, R. (1992). Replication and Consensus: Untouchability, Caste and Ideology in India. *Man*, pages 155–173.
Dinesen, P. T. and Sønderskov, K. M. (2015). Ethnic Diversity and Social Trust. *American Sociological Review*, 80(3):550–573.
Easterly, W. and Levine, R. (1997). Africa's Growth Tragedy: Policies and Ethnic Divisions. *Quarterly Journal of Economics*, 112(4):1203–1250.
Ejdemyr, S., Kramon, E., and Robinson, A. L. (2017). Segregation, Ethnic Favoritism, and the Strategic Targeting of Local Public Goods. *Comparative Political Studies*, pages 1–33.
Fearon, J. D. (2003). Ethnic and Cultural Diversity by Country. *Journal of Economic Growth*, 8(2):195–222.
Fehr, E., Hoff, K., and Kshetramade, M. (2008). Spite and development. *American Economic Review*, 98(2):494–499.
Fotheringham, A. S. and Wong, D. W. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A*, 23(7):1025–1044.
Gao, E. (2016). Tribal Mobilization, Fragmented Groups, and Public Goods Provision in Jordan. *Comparative Political Studies*, 10(49):1372–1403.
Gerring, J., Thacker, S. C., Lu, Y., and Huang, W. (2015). Does Diversity Impair Human Development? A Multi-level Test of the Diversity Debit Hypothesis. *World Development*, 66:166–188.
Gerring, J., Thacker, S. C., Lu, Y., and Oncel, E. (2017). When are Minorities Worse off? A Systematic Investigation of Size and Status. *The Journal of Development Studies*, 53(12):2086–2101.

Gershman, B. and Rivera, D. (2018). Subnational diversity in sub-saharan africa: Insights from a new dataset. *Journal of Development Economics*, 133:231–263.

Gisselquist, R. M., Leiderer, S., and Niño Zarazúa, M. (2016). Ethnic Heterogeneity and Public Goods Provision in Zambia: Evidence of a Subnational "Diversity Dividend". *World Development*, 78:308–323.

Gupta, D. (2005). Caste and Politics: Identity over System. *Annual Review of Anthropology*, 34:409–427.

Guru, G. (2009). Archaeology of Untouchability. *Economic and political weekly*, pages 49–56.

Habyarimana, J., Humphreys, M., Posner, D. N., and Weinstein, J. M. (2007). Why Does Ethnic Diversity Undermine Public Goods Provision? *American Political Science Review*, 101(4):709–725.

Heller, P., Harilal, K., and Chaudhuri, S. (2007). Building Local Democracy: Evaluating the Impact of Decentralization in Kerala, India. *World development*, 35(4):626–648.

Htun, M. (2004). Is Gender like Ethnicity? The Political Representation of Identity Groups. *Perspectives on Politics*, 2(3):439–458.

Jaffrelot, C. (2006). The impact of affirmative action in india: More political than socioeconomic. *India Review*, 5(2):173–189.

Jha, S. (2013). Trade, Institutions, and Ethnic Tolerance: Evidence from South Asia. *American Political Science Review*, 107(4):806–832.

Kasara, K. (2013). Separate and Suspicious: Local Social and Political Context and Ethnic Tolerance in Kenya. *The Journal of Politics*, 75(4):921–936.

Kelly, M. (2019). The Standard Errors of Persistence. *CEPR Discussion Paper*, (DP13783):1–44.

Keshri, K. and Bhagat, R. B. (2012). Temporary and seasonal migration: Regional pattern, characteristics and associated factors. *Economic and Political Weekly*, pages 81–88.

King, G. (2013). *A solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton University Press.

Kone, Z. L., Liu, M. Y., Mattoo, A., Ozden, C., and Sharma, S. (2018). Internal Borders and Migration in India. *Journal of Economic Geography*, 18(4):729–759.

La Porta, R. R., Lopez-de Silanes, F., Shleifer, A., Vishny, R. W., and Vishney, R. (1999). The Quality of Government. *Journal of Law, Economics, and Organization*, 15(1):222–279.

Laitin, D. D. (1989). Language Policy and Political Strategy in India. *Policy Sciences*, 22(3-4):415–436.

Lee, A. (2018). Ethnic Diversity and Ethnic Discrimination: Explaining Local Public Goods Provision. *Comparative Political Studies*, 10(51):1351–1383.

McDonnell, E. M. (2016). Conciliatory states: Elite ethno-demographics and the puzzle of public goods within diverse african states. *Comparative Political Studies*, 49(11):1513–1549.

Miguel, E. and Gugerty, M. K. (2005). Ethnic Diversity, Social Sanctions, and Public Goods in Kenya. *Journal of Public Economics*, 89(11-12):2325–2368.

Montalvo, J. G. and Reynal-Querol, M. (2005). Ethnic Polarization, Potential Conflict, and Civil Wars. *American Economic Review*, 95(3):796–816.

Montalvo, J. G. and Reynal-Querol, M. (2017). Ethnic Diversity and Growth: Revisiting the Evidence . Barcelona Graduate School of Economics Working Paper Series 992.

Mookherjee, D. (2015). Political decentralization. *Annual Review of Economics*, 7(1):231–249.

Munshi, K. and Rosenzweig, M. (2016). Insiders and Outsiders: Local Ethnic Politics and Public Goods Provision . NBER Working Paper No. 21720.

Openshaw, S. (1984). Ecological Fallacies and the Analysis of Areal Census Data. *Environment and Planning A*, 16(1):17–31.

Ostrom, E. (1990). *Governing the Commons*. Cambridge University Press, Cambridge.

Parthasarathy, R. (2017). Ethnic quotas as term-limits: Caste and distributive politics in south india. *Comparative Political Studies*, 50(13):1735–1767.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, 49(1):65–85.

Pierskalla, J. H. and Hollenbach, F. M. (2013). Technology and collective action: The effect of cell phone coverage on political violence in africa. *American Political Science Review*, 107(2):207–224.

Robinson, A. L. (2017). Ethnic Diversity, Segregation and Ethnocentric Trust in Africa. *British Journal of Political Science*, pages 1–23.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15(3):351–357.

Rosenzweig, M. R. and Stark, O. (1989). Consumption Smoothing, Migration, and Marriage: Evidence from Rural India. *Journal of political Economy*, 97(4):905–926.

Saez, L. and Sinha, A. (2010). Political cycles, political institutions and public expenditure in india, 1980–2000. *British Journal of Political Science*, 40(1):91–113.

Sanyal, P. and Rao, V. (2018). *Oral Democracy: Deliberation in Indian Village Assemblies*. Cambridge University Press.

Singh, H. (1994). Constitutional Base for Panchayati Raj in India: The 73rd Amendment Act. *Asian Survey*, 34(9):818–827.

Singh, P. (2015a). *How Solidarity Works for Welfare*. Cambridge University Press.

Singh, P. (2015b). Subnationalism and Social Development: A Comparative Analysis of Indian States. *World Politics*, 67(03):506–562.

Singh, P. and vom Hau, M. (2016). Ethnicity in Time: Politics, History, and the Relationship between Ethnic Diversity and Public Goods Provision. *Comparative Political Studies*, 49(10):1303–1340.

Soifer, H. D. (2016). Regionalism, Ethnic Diversity, and Variation in Public Good Provision by National States. *Comparative Political Studies*, 49(10):1341–1371.

Soifer, H. D. (2019). Units of Analysis in Subnational Research. In *Inside Countries*, pages 92–112. Cambridge University Press.

Tajima, Y., Samphantharak, K., and Ostwald, K. (2018). Ethnic Segregation and Public Goods: Evidence from Indonesia. *American Political Science Review*, 112(3):1–17.

Thachil, T. and Teitelbaum, E. (2015). Ethnic Parties and Public Spending: New Theory and Evidence from the Indian States. *Comparative Political Studies*, 48(11):1389–1420.

Thapliyal, S., Mukherji, A., and Malghan, D. (2019). Economic inequality and loss of commons: Evidence from india. *World Development*, 122:693–712.

Tolnay, S. E. and Beck, E. M. (1995). *A Festival of Violence: An Analysis of Southern Lynchings, 1882-1930*. University of Illinois Press.

Trounstine, J. (2016). Segregation and Inequality in Public Goods. *American Journal of Political Science*, 60(3):709–725.

van der Meer, T. and Tolsma, J. (2014). Ethnic Diversity and Its Effects on Social Cohesion. *Annual Review of Sociology*, 40(1):459–478.

Varshney, A. (2003). *Ethnic Conflict and Civic Life: Hindus and Muslims in India*. Yale University Press.

Wilkinson, S. I. (2006). *Votes and Violence: Electoral Competition and Ethnic Riots in India*. Cambridge University Press.

Wimmer, A. (2015). Is Diversity Detrimental? Ethnic Fractionalization, Public Goods Provision, and the Historical Legacies of Stateness. *Comparative Political Studies*, 49(11):1407–1445.

Table 1. Sample Characteristics: Number of observations

Census Year	**Villages**	**Sub-districts**	**Districts**
2001	590454	5361	583
2011	595983	5878	631

Table 2. Village level incidence of public goods (in %)

	2011	**2001**
Primary School	83	78
Tapwater	45	34
Paved Road	62	53
Electricity	83	76
N	595983	590454

Table 3. Public Goods and Caste Fractionalization

	District	**Sub-District**	**Village**
Primary School	-0.03	-0.07***	0.05***
	(0.04)	(0.01)	(0.00)
Tap Water	-0.16*	0.01	0.03***
	(0.07)	(0.03)	(0.00)
Paved Road	0.04	0.03	0.07***
	(0.06)	(0.02)	(0.00)
Electricity	0.08	0.10***	0.03***
	(0.05)	(0.02)	(0.00)
N	**631**	**5878**	**590051**

Dependent variables for columns (1) and (2): Share of villages with access to different facilities.
Dependent variables for columns (3): Whether a village has access to different facilities.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Control variables for columns (1) and (2): Log Population, Share of SC, ST, Literacy Rate, Log population, number of villages, Sex Ratio, Workforce Participation Rate.
Control variables for column (3): Log Population, Share SC, ST, Literacy Rate, Sex Ratio, Workforce Participation Rate.
Robust standard errors are reported in the parentheses.

Table 4. Change in access to public goods between 2001-2011

	District	Sub-district	Village
Primary School	0.01	-0.05***	0.02***
	(0.03)	(0.01)	(0.00)
Tap Water	-0.11	-0.06*	0.02***
	(0.06)	(0.03)	(0.00)
Paved Road	0.03	0.02	0.04***
	(0.06)	(0.03)	(0.00)
Electricity	-0.02	0.11***	0.03***
	(0.09)	(0.02)	(0.00)
N	582	5179	578,707

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Note 1: The change in the share of villages with access to public goods is the main outcome variable for district and sub-district level regressions.
Note 2: The main outcome variable for the village-level regressions is the change in access to a particular public good. This variable is coded as 0 if there is no change between the two census years and 1 if there has been a positive change. We exclude those villages, here, which lost access to public goods between 2001 and 2011.
Control variables for district & sub-district level models: Log population, share of SC, ST, literacy rate, number of villages, sex ratio, workforce participation rate, and share of villages in the district with respective public goods in 2001.
Control variables for village level models: Log population, share of SC, ST, literacy rate, distance to nearest town, sex ratio, workforce participation rate, 2001 level of access to a given public good, and sub-district fixed effects.
Robust standard errors are reported in the parentheses.

Diversity Deficit and Scale-Flip

(Supplementary Materials)

This document contains all tables referenced in the "Robustness Checks" section of the main paper.

Contents

A Full Regression Results

A.1 Public Goods (2011)

Table A.1.1: Public Goods & Caste Fractionalization (District)

	Primary School	Tap Water	Paved Road	Electricity
Fractionalization	-0.03	-0.16*	0.04	0.08
	(0.05)	(0.08)	(0.07)	(0.07)
Share SC	-0.11	0.11	-0.17	-0.19
	(0.08)	(0.15)	(0.12)	(0.11)
Share ST	-0.02	-0.10	-0.10	-0.10*
	(0.03)	(0.06)	(0.05)	(0.05)
Log population	0.08***	0.04*	0.08***	0.03*
	(0.01)	(0.01)	(0.02)	(0.01)
Number of villages	-0.00***	-0.00***	-0.00***	-0.00**
	(0.00)	(0.00)	(0.00)	(0.00)
Share literate	0.03	0.27**	0.32***	0.36***
	(0.05)	(0.09)	(0.08)	(0.09)
Sex Ratio	-0.07	0.10	0.02	-0.26*
	(0.11)	(0.15)	(0.19)	(0.13)
Workforce participation rate	0.26***	0.20	-0.17	-0.11
	(0.08)	(0.17)	(0.13)	(0.11)
R-Square	0.70	0.84	0.82	0.84
N	631	631	631	631

Dependent variables: Share of villages with access to different facilities.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
All regressions control for state fixed effects.
Robust standard errors are reported in the parentheses.

Table A.1.2: Public Goods & Caste Fractionalization (Sub-district)

	Primary School	Tap Water	Paved Road	Electricity
Fractionalization	-0.07***	0.01	0.03	0.10***
	(0.01)	(0.03)	(0.02)	(0.02)
Share SC	0.06**	-0.07	-0.06	-0.05
	(0.02)	(0.04)	(0.04)	(0.04)
Share ST	0.02*	-0.27***	-0.21***	-0.01
	(0.01)	(0.02)	(0.02)	(0.02)
Log population	0.04***	0.02***	0.06***	0.03***
	(0.00)	(0.00)	(0.00)	(0.00)
Number of villages	-0.00***	-0.00***	-0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
Share literate	0.08***	0.30***	0.36***	0.44***
	(0.02)	(0.04)	(0.04)	(0.03)
Sex Ratio	0.05	0.04	0.04	0.06
	(0.03)	(0.06)	(0.06)	(0.05)
Workforce participation rate	0.04	-0.12*	-0.16***	-0.18***
	(0.03)	(0.05)	(0.05)	(0.04)
R-Square	0.80	0.89	0.82	0.83
N	5878	5878	5878	5878

Dependent variables: Share of villages in a sub-district with access to different facilities.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
All regressions control for district fixed effects.
Robust standard errors are reported in the parentheses.

Table A.1.3: Public Goods & Caste Fractionalization (Village)

	Primary School	Tap Water	Paved Road	Electricity
Fractionalization	0.05***	0.03***	0.07***	0.03***
	(0.00)	(0.00)	(0.00)	(0.00)
Share SC	0.03***	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Share ST	0.06***	-0.07***	-0.07***	-0.04***
	(0.00)	(0.00)	(0.00)	(0.00)
Log population	0.16***	0.07***	0.10***	0.04***
	(0.00)	(0.00)	(0.00)	(0.00)
Area	-0.00***	0.00***	0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
Share literate	0.13***	0.18***	0.22***	0.25***
	(0.00)	(0.00)	(0.00)	(0.00)
Sex Ratio	0.02***	0.01***	0.01*	0.03***
	(0.00)	(0.00)	(0.00)	(0.00)
Workforce participation rate	0.08***	-0.02***	-0.05***	-0.02***
	(0.00)	(0.00)	(0.00)	(0.00)
R-Square	0.40	0.55	0.42	0.61
N	594253	594253	594253	594253

Dependent variables: whether a village has access to different facilities.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
All regressions control for sub-district fixed effects.
Robust standard errors are reported in the parentheses.

A.2 Change in Public Goods (2001 - 2011)

Table A.2.1: Change in access to public goods between 2001-2011 (District)

	Primary School	Tap Water	Paved Road	Electricity
Fractionalization	0.01	-0.11	0.03	-0.02
	(0.03)	(0.06)	(0.06)	(0.08)
Share SC	0.01	0.28*	0.07	0.20
	(0.06)	(0.13)	(0.11)	(0.16)
Share ST	-0.02	0.04	-0.01	0.09
	(0.02)	(0.04)	(0.04)	(0.06)
Log population	0.02	0.06***	0.11***	0.03
	(0.01)	(0.01)	(0.02)	(0.02)
Share literate	-0.08*	0.29***	0.08	0.14
	(0.04)	(0.07)	(0.06)	(0.10)
Sex Ratio	0.02	0.18	0.20	-0.32
	(0.05)	(0.09)	(0.12)	(0.18)
Workforce participation rate	0.05	-0.01	-0.03	-0.28
	(0.06)	(0.11)	(0.10)	(0.16)
Log number of villages	-0.03	-0.06***	-0.11***	-0.03
	(0.02)	(0.01)	(0.02)	(0.02)
R-Square	0.64	0.67	0.78	0.64
N	582	582	582	582

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The change in the share of villages with access to public goods is the main outcome variable.

State fixed effects and initial public goods access have been controlled for in all regressions

Robust standard errors are reported in the parentheses.

Table A.2.2: Change in access to public goods between 2001-2011 (Sub-district)

	Primary School	Tap Water	Paved Road	Electricity
Fractionalization	-0.05**	-0.06	0.02	0.11***
	(0.02)	(0.04)	(0.03)	(0.03)
Share SC	0.01	0.03	-0.03	-0.11**
	(0.02)	(0.06)	(0.05)	(0.04)
Share ST	-0.01	-0.24***	-0.20***	-0.02
	(0.01)	(0.03)	(0.03)	(0.02)
Log population	0.02***	0.01	0.05***	0.01
	(0.00)	(0.01)	(0.01)	(0.01)
Share literate	-0.00	0.14**	0.14***	0.17***
	(0.02)	(0.05)	(0.04)	(0.04)
Sex Ratio	0.02	-0.09	0.01	-0.14
	(0.05)	(0.08)	(0.08)	(0.08)
Workforce Participation Rate	-0.01	0.02	-0.04	-0.11*
	(0.03)	(0.06)	(0.06)	(0.05)
Log number of villages	-0.01***	-0.04***	-0.05***	0.00
	(0.00)	(0.01)	(0.01)	(0.00)
R-Square	0.79	0.76	0.83	0.87
N	5179	5179	5179	5179

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The change in the share of villages with access to public goods is the main outcome variable.

District fixed effects and initial public goods access have been controlled for in all regressions

Robust standard errors are reported in the parentheses.

B Loss of Public Goods

Table B.1: Change in access to public goods between 2001-2011 (Village)

	Primary School	Tap Water	Paved Road	Electricity
Fractionalization	0.02***	0.02***	0.04***	0.03***
	(0.00)	(0.00)	(0.00)	(0.00)
Share SC	0.02***	-0.01***	-0.01*	-0.02***
	(0.00)	(0.00)	(0.00)	(0.00)
Share ST	0.04***	-0.06***	-0.05***	-0.05***
	(0.00)	(0.00)	(0.00)	(0.00)
Log population	0.05***	0.04***	0.04***	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)
Sex Ratio	0.01*	0.00	0.00	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Share cultivator	0.02***	-0.03***	-0.02***	-0.00**
	(0.00)	(0.00)	(0.00)	(0.00)
Log distance to nearest town	0.00***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
R-Square	0.45	0.39	0.49	0.71
N	578707	578707	578707	578707

Dependent variables: whether a village gained access to a given public good between 2001 and 2011.
Each variable takes a value of 1 if a village gained access to a given public good, and 0 if there was no change in access to the same public good or the village lost access to the public good.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
All regressions control for sub-district fixed effects and initial public goods access (2001).
Robust standard errors are reported in the parentheses.

Table B.2: Fractionalization and Loss of Public Goods (Village)

	Primary School	Tap Water	Paved Road	Electricity
Fractionalization	-0.02***	0.00	-0.02***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Share SC	-0.00*	0.00	0.01***	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Share ST	-0.02***	0.03***	0.04***	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Log population	-0.05***	-0.03***	-0.06***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Sex Ratio	-0.01**	-0.01***	-0.00	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Share cultivators	-0.01***	0.02***	0.03***	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Log distance to nearest town	-0.00***	-0.00	0.01***	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
R-Square	0.18	0.49	0.36	0.71
N	525884	475072	452811	494372

Dependent variables: whether a village lost access to a given public good between 2001 and 2011.
Each variable takes a value of 1 if a village lost access to a given public good, and 0 if there was no change in access to the same public good between 2001 and 2011.
We exclude those villages in this sub-sample which gained access to a given public good between years 2001 and 2011.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
All regressions control for sub-district fixed effects and initial public goods access (2001).
Robust standard errors are reported in the parentheses.

C Other Backward Castes

Table C.1: Caste-Fractionalization & Public Goods: Bottom and Top OBC Share (District)

	Bottom Quartile Districts				Top Quartile Districts			
	Primary School	Tap Water	Paved Road	Electricity	Primary School	Tap Water	Paved Road	Electricity
Fractionalization	-0.08	-0.14	-0.21	0.12	-0.34	-1.35	-1.06*	0.07
	(0.09)	(0.13)	(0.12)	(0.10)	(0.23)	(0.95)	(0.43)	(0.32)
Share SC	-0.20	-0.03	0.21	-0.43**	0.49	1.48	1.19*	0.18
	(0.13)	(0.32)	(0.27)	(0.15)	(0.30)	(1.09)	(0.52)	(0.36)
Share ST	0.06	-0.07	-0.14	-0.27**	0.32	1.66	1.22**	-0.01
	(0.07)	(0.10)	(0.10)	(0.10)	(0.24)	(1.18)	(0.40)	(0.35)
Log population	0.07**	0.02	0.03	0.03	0.10***	0.06*	0.11**	0.02
	(0.02)	(0.03)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)
Number of villages	-0.00***	-0.00*	-0.00**	-0.00*	-0.00***	-0.00*	-0.00**	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Share literate	-0.07	0.15	0.50**	0.33	-0.06	0.49*	0.35	0.53**
	(0.13)	(0.21)	(0.17)	(0.20)	(0.12)	(0.19)	(0.23)	(0.19)
Sex Ratio	0.06	-0.02	-0.72	-0.12	-0.17	-0.24	-0.06	-0.55
	(0.28)	(0.27)	(0.44)	(0.28)	(0.24)	(0.35)	(0.51)	(0.33)
Workforce participation rate	0.51**	0.31	-0.16	0.74**	0.04	0.14	-0.00	-0.37
	(0.17)	(0.42)	(0.25)	(0.27)	(0.13)	(0.27)	(0.25)	(0.21)
R-Square	0.78	0.82	0.82	0.75	0.74	0.91	0.75	0.90
N	158				156			

Dependent variables: Share of villages with access to different facilities.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
All regressions control for state fixed effects.
Robust standard errors are reported in the parentheses.
OBC population shares for districts have been computed using NSSO 68th round (2011-12).

Table C.2: Caste-Fractionalization & Public Goods: Bottom and Top OBC Share (Sub-district)

	Bottom Quartile Districts				Top Quartile Districts			
	Primary School	Tap Water	Paved Road	Electricity	Primary School	Tap Water	Paved Road	Electricity
Fractionalization	-0.09**	0.03	0.08	0.16**	-0.08*	-0.09	0.02	0.04
	(0.03)	(0.06)	(0.06)	(0.05)	(0.03)	(0.10)	(0.11)	(0.06)
Share SC	0.08*	-0.11	-0.19*	-0.15**	0.06	-0.07	-0.05	0.06
	(0.04)	(0.09)	(0.08)	(0.05)	(0.05)	(0.13)	(0.14)	(0.10)
Share ST	0.06*	-0.09	-0.11*	0.02	0.04*	-0.35***	-0.23*	0.06
	(0.02)	(0.05)	(0.04)	(0.03)	(0.02)	(0.08)	(0.10)	(0.04)
Log population	0.03***	-0.00	0.04**	0.04***	0.04***	0.01	0.08***	0.03*
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Number of villages	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00**	-0.00***	-0.00**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Share literate	0.09	0.38***	0.36***	0.47***	0.06	0.41***	0.41***	0.53***
	(0.04)	(0.09)	(0.08)	(0.07)	(0.04)	(0.10)	(0.11)	(0.12)
Sex Ratio	0.04	0.04	-0.06	0.12	-0.07	0.09	0.07	-0.06
	(0.08)	(0.12)	(0.11)	(0.13)	(0.07)	(0.16)	(0.14)	(0.14)
Workforce participation rate	0.04	-0.19	-0.13	-0.12	0.03	-0.03	0.04	-0.11
	(0.06)	(0.13)	(0.10)	(0.11)	(0.05)	(0.09)	(0.12)	(0.12)
R-Square	0.83	0.81	0.80	0.79	0.75	0.93	0.79	0.85
N		1473				1467		

Dependent variables: Share of villages with access to different facilities.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
All regressions control for district fixed effects.
Robust standard errors are reported in the parentheses.
OBC population shares for districts have been computed using NSSO 68th round (2011-12).

Table C.3: Caste-Fractionalization & Public Goods: Bottom and Top OBC Share (Village)

	Bottom Quartile Districts				Top Quartile Districts			
	Primary School	Tap Water	Paved Road	Electricity	Primary School	Tap Water	Paved Road	Electricity
Fractionalization	0.04***	0.03***	0.08***	0.01***	0.10***	0.04***	0.06***	0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
Share SC	0.03***	-0.03***	-0.04***	0.00	0.01	-0.00	0.01*	0.01**
	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
Share ST	0.06***	-0.06***	-0.05***	-0.02***	0.06***	-0.09***	-0.10***	-0.01
	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
Log population	0.16***	0.06***	0.10***	0.03***	0.17***	0.05***	0.08***	0.04***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of villages	-0.00***	0.00*	0.00	-0.00***	-0.00***	0.00***	0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Share literate	0.10***	0.23***	0.26***	0.28***	0.13***	0.16***	0.19***	0.17***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sex Ratio	0.01	0.01**	-0.01*	0.01***	0.02**	0.02*	0.02**	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Workforce participation rate	0.05***	-0.01	-0.07***	-0.02***	0.10***	-0.02*	-0.03**	-0.01
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
R-Square	0.41	0.58	0.40	0.60	0.41	0.53	0.44	0.69
N	149912				148940			

Dependent variables: whether a village has access to different facilities.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
All regressions control for sub-district fixed effects.
Robust standard errors are reported in the parentheses.

D Temporal Stability

Table D.1: SC/ST Proportion, 2001 - 2011 (Rural residents only)

Correlation between π_{01}^{SCST} and π_{11}^{SCST}		
District	0.9846	N = 582
Sub-district	0.9813	N = 5,272
Village	0.9468	N = 578,889

Table D.2: Fractionalization, 2001 - 2011 (Rural residents only)

Correlation between FRA_{01} and FRA_{02}		
District	0.9575	N = 582
Sub-district	0.9501	N = 5,272
Village	0.9028	N = 578,889

E Marginalized Groups

Table E.1: Public Goods & Interaction between Fractionalization and Proportion SC-ST (District)

	Primary School	Tap Water	Paved Road	Electricity
Fractionalization	-0.02	-0.06*	-0.02	0.03
	(0.01)	(0.02)	(0.02)	(0.02)
Fractionalization X Share SCST	0.02	0.03	0.09**	-0.02
	(0.02)	(0.03)	(0.03)	(0.03)
Share SC	-0.12	0.09	-0.22	-0.18
	(0.09)	(0.16)	(0.12)	(0.11)
Share ST	-0.06	-0.15	-0.26**	-0.06
	(0.06)	(0.08)	(0.09)	(0.08)
Log population	0.08***	0.04*	0.08***	0.03*
	(0.01)	(0.01)	(0.02)	(0.01)
Number of villages	-0.00***	-0.00***	-0.00***	-0.00**
	(0.00)	(0.00)	(0.00)	(0.00)
Share literate	0.03	0.28**	0.34***	0.35***
	(0.05)	(0.09)	(0.09)	(0.09)
Sex ratio	-0.08	0.09	-0.03	-0.25
	(0.12)	(0.15)	(0.19)	(0.13)
Workforce participation rate	0.27***	0.2	-0.16	-0.12
	(0.08)	(0.17)	(0.13)	(0.11)
R-Square	0.7	0.84	0.82	0.84
N		631		

Dependent variables: Share of villages with access to different facilities.
The two main variables: fractionalization and proportion SC-ST have been standardized.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
All regressions control for state fixed effects.
Robust standard errors are reported in the parentheses.

Table E.2: Public Goods & Interaction between Fractionalization and Proportion SC-ST (Sub-district)

	Primary School	Tap Water	Paved Road	Electricity
Fractionalization	-0.02**	0	-0.01	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)
FractionalizationXShareSCST	-0.01	0	0.04**	0.01
	(0.01)	(0.02)	(0.01)	(0.01)
Share SC	0.07**	-0.07	-0.10*	-0.06
	(0.02)	(0.04)	(0.04)	(0.04)
Share ST	0.04**	-0.27***	-0.28***	-0.03
	(0.01)	(0.03)	(0.03)	(0.02)
Log population	0.04***	0.02***	0.06***	0.03***
	(0.00)	(0.00)	(0.00)	(0.00)
Number of villages	-0.00***	-0.00***	-0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
Share literate	0.08***	0.30***	0.36***	0.44***
	(0.02)	(0.04)	(0.04)	(0.03)
Sex ratio	0.04	0.04	0.05	0.06
	(0.03)	(0.06)	(0.06)	(0.05)
Workforce participation rate	0.04	-0.12*	-0.16***	-0.18***
	(0.03)	(0.05)	(0.05)	(0.04)
R-Square	0.8	0.89	0.82	0.83
N	5878			

Dependent variables: Share of villages in a sub-district with access to different facilities.
The two main variables: fractionalization and proportion SC-ST have been standardized.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
All regressions control for district fixed effects.
Robust standard errors are reported in the parentheses.

Table E.3: Public Goods & Interaction between Fractionalization and Proportion SC-ST (Village)

	Primary School	Tap Water	Paved Road	Electricity
Fractionalization	0.02***	0.01***	0.03***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
FractionalizationXShareSCST	-0.01***	-0.01*	0.00	0.04**
	(0.00)	(0.00)	(0.00)	(0.00)
Share SC	0.03***	0	0	-0.02***
	(0.00)	(0.00)	(0.00)	(0.00)
Share ST	0.07***	-0.07***	-0.08***	-0.06***
	(0.00)	(0.00)	(0.00)	(0.00)
Log population	0.16***	0.07***	0.10***	0.04***
	(0.00)	(0.00)	(0.00)	(0.00)
Area	-0.00***	0.00**	0.00***	-0.00**
	(0.00)	(0.00)	(0.00)	(0.00)
Share literate	0.13***	0.18***	0.22***	0.25***
	(0.00)	(0.00)	(0.01)	(0.00)
Sex Ratio	0.02***	0.01***	0.01*	0.03***
	(0.00)	(0.00)	(0.00)	(0.00)
Workforce Participation Rate	0.08***	-0.02***	-0.05***	-0.02***
	(0.00)	(0.00)	(0.00)	(0.00)
R-Square	0.4	0.55	0.42	0.61
N	594253			

Dependent variables: whether a village has access to different facilities.
The two main variables: fractionalization and proportion SC-ST have been standardized.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
All regressions control for sub-district fixed effects.
Robust standard errors are reported in the parentheses.

Table E.4: Fractionalization and Public Goods: SC-ST Proportion Partition (District)

	Share SC-ST ≥ 0.5				Share SC=ST < 0.5			
	Primary School	Tap Water	Paved Road	Electricity	Primary School	Tap Water	Paved Road	Electricity
Fractionalization	-0.22	-0.41	-0.13	-0.25	-0.20	-0.23	-0.69**	0.26
	(0.19)	(0.31)	(0.26)	(0.36)	(0.20)	(0.31)	(0.23)	(0.21)
Share SC	-0.02	-0.46	-0.65	-1.60	0.13	0.20	0.75**	-0.27
	(0.29)	(0.40)	(0.55)	(0.86)	(0.23)	(0.36)	(0.26)	(0.23)
Share ST	-0.05	-0.54	-0.04	-0.62	0.15	-0.04	0.71**	-0.32
	(0.20)	(0.29)	(0.33)	(0.41)	(0.21)	(0.34)	(0.25)	(0.26)
Log population	0.13***	0.03	0.08	0.08*	0.06***	0.04**	0.08***	0.02
	(0.02)	(0.04)	(0.06)	(0.03)	(0.01)	(0.01)	(0.02)	(0.01)
Number of villages	-0.00***	-0.00*	-0.00	-0.00**	-0.00***	-0.00***	-0.00***	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Share literate	0.14	0.15	0.48**	0.35	-0.02	0.31**	0.33***	0.35***
	(0.13)	(0.21)	(0.17)	(0.22)	(0.06)	(0.12)	(0.09)	(0.08)
Sex Ratio	-0.18	-0.03	-1.44*	-0.70	-0.18	0.09	-0.09	-0.13
	(0.40)	(0.53)	(0.71)	(0.61)	(0.14)	(0.17)	(0.20)	(0.10)
Workforce participation rate	0.28	-0.29	-0.10	0.50	0.19*	0.19	-0.40**	-0.29**
	(0.21)	(0.48)	(0.30)	(0.33)	(0.09)	(0.18)	(0.14)	(0.10)
R-Square	0.79	0.79	0.83	0.71	0.68	0.86	0.83	0.90
N	129				502			

Dependent variables: Share of villages with access to different facilities.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
All regressions control for state fixed effects.
Robust standard errors are reported in the parentheses.

Table E.5: Fractionalization and Public Goods: SC-ST Proportion Partition (Sub-district)

	Share SC-ST ≥ 0.5				Share SC-ST < 0.5			
	Primary School	Tap Water	Paved Road	Electricity	Primary School	Tap Water	Paved Road	Electricity
Fractionalization	-0.13	-0.11	-0.01	0.12	-0.09	-0.26*	-0.15	-0.00
	(0.07)	(0.13)	(0.13)	(0.13)	(0.07)	(0.12)	(0.11)	(0.09)
Share SC	0.04	-0.24	-0.16	0.12	0.10	0.24	0.10	0.04
	(0.08)	(0.13)	(0.17)	(0.16)	(0.07)	(0.14)	(0.12)	(0.09)
Share ST	0.01	-0.28*	-0.19	0.05	0.06	0.18	0.03	0.06
	(0.07)	(0.11)	(0.14)	(0.13)	(0.08)	(0.14)	(0.12)	(0.10)
Log population	0.03***	0.03	0.07***	0.04**	0.05***	0.01*	0.06***	0.02***
	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Number of villages	-0.00***	-0.00***	-0.00***	-0.00**	-0.00***	-0.00***	-0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Share literate	0.15**	0.24*	0.61***	0.71***	0.03	0.35***	0.26***	0.30***
	(0.06)	(0.11)	(0.10)	(0.11)	(0.03)	(0.06)	(0.05)	(0.04)
Sex Ratio	0.08	-0.10	-0.12	0.11	0.05	0.06	0.20**	0.06
	(0.10)	(0.18)	(0.16)	(0.16)	(0.06)	(0.07)	(0.08)	(0.06)
Workforce participation rate	0.01	-0.17	0.06	-0.41**	0.06	-0.05	-0.21**	-0.14**
	(0.07)	(0.15)	(0.12)	(0.15)	(0.04)	(0.05)	(0.06)	(0.04)
R-Square	0.83	0.81	0.81	0.77	0.80	0.91	0.83	0.87
N	1244				4634			

Dependent variables: Share of villages in a sub-district with access to different facilities.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
All regressions control for district fixed effects.
Robust standard errors are reported in the parentheses.

Table E.6: Fractionalization and Public Goods: SC-ST Proportion Partition (Village)

	Share SC-ST ≥ 0.5				Share SC-ST < 0.5			
	Primary School	Tap Water	Paved Road	Electricity	Primary School	Tap Water	Paved Road	Electricity
Frationalization	0.01	0.01	0.10***	0.09***	0.13***	0.06***	0.15***	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Share SC	0.03**	-0.05***	0.03*	0.08***	-0.03	-0.05*	-0.11***	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Share ST	0.07***	-0.08***	-0.02	0.04***	-0.04**	-0.09***	-0.18***	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Log population	0.18***	0.06***	0.09***	0.05***	0.16***	0.07***	0.09***	0.04***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Area	-0.00***	0.00**	0.00***	-0.00**	-0.00***	0.00***	0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Share literate	0.16***	0.14***	0.23***	0.31***	0.10***	0.21***	0.22***	0.20***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sex Ratio	0.01	-0.00	0.01*	0.01**	0.02***	0.02***	0.00	0.03***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Workforce Participation Rate	0.07***	-0.00	-0.05***	-0.04***	0.08***	-0.04***	-0.05***	-0.01*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
R-Square	0.42	0.52	0.45	0.58	0.41	0.55	0.41	0.64
N	174795				419458			

Dependent variables: whether a village has access to different facilities.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
All regressions control for sub-district fixed effects.
Robust standard errors are reported in the parentheses.