Polity size and local government performance: evidence from India*

Veda Narasimhan† and Jeffrey Weaver‡

Abstract

Developing countries have increasingly decentralized power to local governments. This paper studies a central element of decentralization – polity size – using population-based discontinuities that determine local government boundaries for over 100,000 Indian villages. Over the short and long-run, individuals allocated into local governments with smaller populations have better access to public goods. However, this relationship is non-linear, with benefits only observed below certain population thresholds. We provide evidence that these results are explained by greater civic engagement and improved leader selection in those jurisdictions. We find no evidence for other theorized mechanisms such as elite capture.

Keywords: local government, decentralization, gram panchayat, India
JEL codes: H41, O12, H75, P43, D72

*We are grateful for helpful comments from Vittorio Bassi, David Blakeslee, Bruno Caprettini, Lorenzo Casaburi, Matthew Kahn, Guilherme Lichand, Paul Niehaus, Paulina Oliva, Joachim Voth, and David Yanagizawa-Drott, as well as seminar audiences at American University, Dartmouth, ETH Zurich, NBER Summer Institute (Development), NYU - Abu Dhabi, PacDev, Stockholm University, University of British Columbia, University of Southern California, University of St. Gallen, University of Washington, University of Zurich, and Yale University. All views expressed are those of the authors. Funding for this project was provided by the Swiss National Science Foundation (grant no. 187942), and the Forschungskredit of the University of Zurich (grant no. FK-19-017). This paper subsumes and replaces our previous work in Narasimhan (2021) and Weaver (2021).

†Department of Economics, New York University - Abu Dhabi, veda.narasimhan@nyu.edu
‡Department of Economics, University of Southern California, jbweaver@usc.edu.

1 Introduction

Developing countries have increasingly moved towards political decentralization and devolution of power to elected local governments. One key aspect of decentralization is the division of the population into local government units, where the size of these polities may be an important determinant of their performance. A number of models suggest decentralization generates benefits by dividing a population into a larger number of smaller local governments. This may better cater to heterogeneous tastes (Oates, 1972) or incentivize public good provision through competition and yardstick comparisons (Tiebout, 1956; Besley and Case, 1995). Smaller jurisdictions are likely to be more homogeneous, which may reduce conflict (Mansuri and Rao, 2012; Bazzi and Gudgeon, 2021), and may make it easier to monitor leaders and hold them accountable (Seabright, 1996; Boffa et al., 2016).

Alternatively, there are many potential drawbacks to smaller polities. Smaller jurisdictions may be more easily captured by elites (Ambedkar, 1932; Bardhan and Mookherjee, 2000; Bardhan, 2002), miss out on economies of scale (Alesina and Spolaore, 1997) or fail to internalize externalities (Lipscomb and Mobarak, 2016). A larger unit will also have a larger pool of talent for political leadership and may have more political competition that puts greater pressure on leaders to perform (Gerring et al., 2015). Given the many plausible theoretical mechanisms pushing in different directions, it is unclear how polity size will affect government performance. Understanding these consequences is important for policy as well as for evaluating the empirical relevance of different theories of decentralization.

This paper provides causal evidence on the consequences of polity size and tests of prominent theories of decentralization in the context of the Indian state of Uttar Pradesh. Uttar Pradesh has a population of 241 million people and would be the fifth most populous country in the world if it were an independent country. As in the rest of India, the village council, or gram panchayat (GP), is the lowest level of elected government. Each GP typically contains multiple villages, with local representatives elected from each village and one head (*pradhan)* elected for the entire GP. The GP oversees the provision of local public goods such as water and sanitation, is the primary implementer of a large workfare program, the National Rural Employment Guarantee Scheme (NREGS), and plays a role in many welfare programs. Outside of those formal responsibilities, GP leaders play a crucial role in lobbying bureaucrats and politicians for the allocation of resource to their GPs.

Our empirical approach takes advantage of a unique rule in Uttar Pradesh that determines the allocation of villages to local governments. This rule stipulates that based on the most recent decadal census, villages whose population exceeds one thousand should be allocated into their own GP. Uttar Pradesh has conducted two rounds of panchayat delimitation based on this rule – once in 1995 and a second time in 2015 – and boundaries have otherwise remained fixed, even if a village's population exceeded the threshold at some time between 1995 and 2015. The rule causes villages with populations just below 1000 to end up in much more populous GPs than villages with populations just above the cutoff; in the 2015 delimitation, villages just below the cutoff ended up in GPs with an average population that is roughly 40% larger than villages above the cutoff. We use this in a regression discontinuity (RD) design to estimate the effect of allocation into a smaller GP.

We combine census and administrative data sets spanning a 28-year period to evaluate the effect of this allocation on public service delivery, including novel village-level data on public benefit receipt. We first investigate village-level infrastructure, and find that villages allocated to smaller GPs have significantly better access to primary and middle school facilities in both the short and long-run, which results in economically meaningful gains in educational attainment. There are also increases in all-weather roads and Fair Price Shops (which provide access to subsidized food), but we do not see any effects on electrification.

Next, we examine welfare programs. Households in villages allocated to smaller polities are more likely to have houses made out of brick rather than organic materials, toilets, and closed drainage systems. These effects are sizable – for example, there is a 17% increase in likelihood of having a toilet. This is explained by improved implementation of programs that fund housing and sanitation improvements, for which local leaders play a substantial role. We also find elevated access to eight benefit programs ranging from pensions to government-provided health insurance, pointing to leaders in smaller polities putting more effort into helping citizens make claims on the state.

Finally, we use administrative data to measure the implementation of the NREGS workfare program. Individuals in villages allocated to smaller GPs have better workfare access, with effects such as NREGS earnings that are around fifteen percent higher. This data also disaggregates program participation by caste, which we use to show that effects are similar across advantaged and disadvantaged groups.

These estimates pertain to the treatment effect of shifting citizens into GPs with approximately 1000 people. However, policymakers may be interested in other polity sizes (e.g. shifting citizens from a GP of 5000 people to one of 4000). We leverage the same delimitation rule to estimate the treatment effects of other reforms. For the intuition of this exercise, suppose that villages a and b were in a GP with village c, but village c's population is around the population cutoff at the time of delimitation. If village c is above the cutoff, it will be split off and villages a and b will remain together in a smaller GP; if below, then villages a and b will remain with it in a larger GP. Using this idea for an alternative RD approach – with village c's population as a running variable for villages a and b – we find evidence of non-linearities, where effects are concentrated in cases where the created GPs have populations of 1000 to 2000. This suggests that the real effect is from creating polities of certain sizes rather than generically making them smaller.

We conduct a large number of robustness checks to show that other government programs do not use the same population cutoff and that our estimates are not the result of mechanical funding rules favoring smaller polities. We additionally document a lack of negative spillovers onto untreated villages and present evidence that this is due to soft budget constraints.

To better understand the drivers of the effects we find, we test mechanisms from the theoretical decentralization literature related to polity size. These tests are helpful in interpreting why there is an increase in service delivery outcomes, as well as being of intrinsic interest in providing evidence on mechanisms that have previously been difficult to document empirically. Political channels appear to play the biggest role. Using the main RD approach and two rounds of GP elections data, we show that civic engagement is greater (Mookherjee, 2015), and that candidates and elected representatives are less likely to have a criminal record in smaller polities. Both effects are consistent with better performance (Seabright, 1996; Boffa et al., 2016). We additionally find no evidence of greater levels of elite capture (Bardhan and Mookherjee, 2000), where high caste and wealthy individuals are no more likely to be elected, and increases in service access are similar across high and low caste groups. Finally, the effects are strongest in villages where delimitation into a smaller GP allows them to elect a leader from within the village, who may have better local information or greater motivation to perform; for villages that already were the majority of the population in their GP, and so could have elected a local leader even if not split off, the discontinuity has no effect.[1]

These results have significant implications for decentralization policy and provide guidance to policymakers on creation of local governments. Most directly, these findings support reductions in local government size in India, where average GP population ranges to over ten thousand people in some states (Figure A1). They also provide principles for the creation of smaller local government bodies across a broader global context, pointing to specific size ranges within which beneficial political mechanisms are activated. Policies reforming polity size are especially exciting since they are straightforward to implement at scale even in the context of low capacity states, as seen here.

The paper contributes to three main strands of the literature. First, we provide novel estimates on the causal effect of polity size on government performance. Work on optimal

[1] We also check whether the observed effects can be explained by creating more homogeneous or geographically compact GPs. We find no evidence for either channel, as well as others such as a higher ratio of leaders to citizens or reductions in interjurisdictional competition.

government scale has been predominantly theoretical due to the difficulty of finding plausibly exogenous sources of variation in the size of political units (Alesina and Spolaore, 1997; Boffa et al., 2016). The most closely related empirical work uses difference-in-differences to study the consequences of splits or amalgamation of municipalities or districts (e.g. Lassen and Serritzlew 2011; Pierskalla 2016; Dahis and Szerman 2021). Our approach expands upon that approach by measuring not only the average effects of amalgamation or fragmentation, but the consequences of specific polity sizes. This provides more implementable guidance for local government creation, especially since one might expect non-linearities in the polity size-performance relationship such that average effects are not necessarily informative; e.g., it cannot be the case that more splits are always better, where jurisdictions that each contain only a single person are almost certainly sub-optimal.[2] Our findings provide lessons that can generalize to settings with different population structures than north India – in particular, the value of creating local governments of a size such that leaders come from the community they represent.

Second, we complement the theoretical literature on political decentralization by examining the empirical relevance of theoretical mechanisms through which polity size may affect outcomes. Polity size is a key element of many of the most prominent theories of decentralization, but finding plausible sources of variation to test these theories has traditionally been challenging. We find evidence of improved accountability and selection into candidacy within smaller local governments. This is consistent with Seabright (1996) and Cremer et al. (1994) among others, who argue that decentralization helps to ease some of the issues associated with political accountability in democracies. We do not find evidence that many plausible mechanisms explain our results, such as a higher ratio of politicians to citizens, elite capture, or lower identity group fractionalization (Alesina et al., 1999; Bazzi and Gudgeon, 2021).

Third, we contribute to the literature linking design of local government to public goods provision in developing countries. A growing literature has explored aspects such as democratization (Olken, 2010; Martinez-Bravo et al., 2022), representative identity (Chattopadhyay and Duflo, 2004; Martinez-Bravo, 2014), and technological solutions to governance problems (Lewis-Faupel et al., 2016; Banerjee et al., 2020). We complement these papers by studying a different aspect in the design of local government. A nice feature of polity size reforms is that they require minimal state capacity to implement, as evidenced by these effects in one

[2] The RD approach likely also provides more generalizable estimates by studying changes induced by an exogenous cutoff. Difference-in-differences estimates are typically based around politically-motivated splits/amalgamations, and so naturally may be limited to a subset of change in which splits are beneficial (Pierskalla, 2016). Previous papers have also focused on much larger levels of aggregation such as districts in Indonesia.

of the lowest capacity states in India.

The remainder of the paper is organized as follows. Section 2 discusses the institutional background, while section 3 describes the data. Section 4 explains the empirical approach, section 5 provides results, and section 6 investigates mechanisms. Section 7 concludes and discusses the policy implications.

2 Background

2.1 History of local governance in India

Local rural village councils - known as gram panchayats (GPs) - have existed in some form in India since the pre-modern era. The first legislation concerning their administration was introduced soon after independence, but it took until the 1990s for a consistent national level policy to be established. Throughout the post-independence period, GPs were given limited responsibilities, and decisions were primarily made at the level of the block (consisting of 50-100 villages) or higher levels or government.

In the 1990s, India and many other large developing countries like China and Indonesia engaged in large-scale devolution of power to elected local governments. In India, this was based on the 73^{rd} Amendment to the Indian Constitution, a comprehensive decentralization reform with three key components: 1) formation of a three-tier system of governance at the village (gram panchayat), sub-district and district levels; 2) devolution of power and responsibility to the gram panchayat; and 3) standardization of political elections at the village-level. By formally amending the constitution, the federal government ensured uniformity in these key aspects, but allowed states discretion over which responsibilities to decentralize and how to form councils.

The 73^{rd} amendment lists 29 potential areas for devolution to local governments (see Table A1), but let individual states decide which to decentralize (Chaudhary and Iyer, 2022). The topics range from welfare programs to economic development (e.g. agriculture, irrigation). Central and state government programs also often use GPs as the point of contact for information dissemination and program administration: for example, over 99% of households who received subsidies for toilet construction in Uttar Pradesh reported getting assistance from GP leaders (Planning Commission of India, 2013). Over time, the roles of the GP have expanded to administering the increasing number of government programs targeted at rural areas, one of the most important expansions being the National Rural Employment Guarantee Scheme (NREGS), a large workfare program.

Leadership of the local government body consists of ward councilors and a head (*prad-*

han). "Wards" typically consist of a few hundred voters, and each elect a single local representative ("ward councilor"). The system for electing pradhans varies from state to state: in the state we focus on, there is direct election by citizens at the same time as the ward councilor positions. For both types of positions, candidates are not officially aligned with political parties on the ballot, the candidate with the most votes wins regardless of whether they have a majority, and elections occur every five years with no formal term limits. However, in each election cycle, a third of positions are reserved for women, and as many as half may be reserved for individuals from traditionally disadvantaged identity groups, which often imposes de facto term limits and thus may affect performance (Dal Bó and Rossi, 2011).

Different states have taken varied approaches in creating GPs. Average GP population across India is just under 3500 people, but this varies substantially. At one extreme, gram panchayats in Bihar, Kerala, and West Bengal are composed of many villages and have average populations of over 10,000 (Figure A1). On the other end, GPs in other states contain an average of less than two thousand people. These choices are largely a function of historical path dependence – for example, the key rule governing GP size in Uttar Pradesh originates in a law from 1947 – and so may not reflect currently optimal policy.

2.2 Local governance in Uttar Pradesh

This paper will focus on village-level governments in Uttar Pradesh, the most populous state in India. Uttar Pradesh is the third poorest state in India, with 38% of its population deemed poor, and has a PPP-adjusted GDP per capita comparable to that of Mali (NITI Aayog, 2021). Nearly four-fifths of its population lives in rural areas (Census of India, 2011), where public service delivery is among the worst in the country.

Uttar Pradesh was one of the first states in India to introduce panchayats under the Uttar Pradesh Panchayat Raj Act of 1947. However, until the 73rd amendment in 1993, the panchayat system was widely regarded as non-functional (Alsop et al., 2000). At that point, Uttar Pradesh amended existing laws to fall in line with the constitutional amendment and devolved significant functionalities to GPs, including education, drinking water, welfare and child development, health, agriculture, and village development.

We focus on Uttar Pradesh since it uses explicit population thresholds to allocate villages to local government units. GP boundaries are determined by officials of the State Election Commission, who join proximate villages to create local government units (Jha, 2021). A typical GP includes 1-4 villages, each of which are themselves a collection of geographically proximate smaller clusters of households called "hamlets". Section 11-F of the 1994 Uttar

Pradesh Panchayat Laws Amendment states that the government will declare "a village or group of villages, having, so far as practicable, a population of one thousand, to be a Panchayat area". It further defines villages as those recorded in the census of India, states that these census-defined villages should not be divided, and specifies use of population counts from the most recent census. An official following this rule will be more likely to allocate villages with a population of 1000 or more into a GP of their own, whereas villages below the cut-off will be put in a GP with other villages. Section 4 will show that officials follow this rule, which we will use to examine the effect of GP size.

Uttar Pradesh has undergone two major rounds of panchayat delimitation. The first occurred in 1995 and was based on data from the 1991 census, leading to the creation of 58,620 gram panchayats. There was a consolidation in 2000 to 52,929 GPs, and the boundaries remained steady for the next fifteen years, with 52,001 GPs in the 2005 elections and 51,914 GPs in 2010. In 2015, there was a second major round of delimitation based on data from the 2011 census, and the number of GPs expanded to 59,074 (Uttar Pradesh State Election Commission, 2021). For the remainder of the paper, we will analyze the 1995-2014 and post-2015 periods as distinct eras with fixed GPs throughout.

2.3 Financing and allocation of public programs

This paper measures how polity size affects public service delivery. To understand these estimates, one must consider the underlying funding and allocation processes. These processes differ across programs – for example, they are different for workfare programs than for school construction – but conceptually, can be divided into three categories.

The first category is services for which a block or district-level bureaucrat determines allocations, and funding comes from the state or central government. A typical example would be construction of a primary school, which is funded by the state or central government (e.g. the District Primary Education Program studied in Khanna (2022)), but district-level bureaucrats decide where to build schools. Although these decisions are technically in the hands of bureaucrats, they are well-known to be influenced by political considerations (Wilkinson, 2006; Vaishnav and Sircar, 2012). As a result, the presence of these types of public goods in a community reflect the effort of local government leaders in lobbying bureaucrats or higher level politicians.

The second category consists of programs that are initiated or implemented by individuals or local governments, but funded by higher levels of government. This includes programs we will study that fund the construction of improved housing and toilets, and the National Rural Employment Guarantee Scheme workfare program, where GP leaders propose projects, hire workers, and generally manage implementation. In these cases, funding decisions must be approved bureaucratically (see Banerjee et al. (2020) for a description of processes for NREGS), but the effort of GP leaders determines awareness of programs as well as what is proposed and successfully implemented.

The final category is services initiated by the local government and funded by its budget. Although gram panchayats have the power to levy taxes on certain economic activities, formal tax collection is minimal at the local level as in many other developing countries; across India, local taxes are 0.4% of total tax revenues (Rao, 2007). Their budget predominantly comes from transfers from the state and central governments, which both are determined by a funding formula from the State Finance Commission (Sethi et al., 2004). Over most of the study period, the formula was linear in GP population (weight of 80%) and the area of the GP (20% weight); in 2011, this switched to 80% based on population and 20% based on SC/ST population share (Chakraborty et al., 2018). Under these formulas, all GPs receive similar per capita transfer amounts regardless of their size; we will later show that these formulas are followed. Over the period 2013-2020, local government budgets in Uttar Pradesh averaged Rs 544 (US$8) per inhabitant.[3] Centre for Policy Research (2019) analyses GP budgets across many states and finds that the largest expenditure categories are the creation of "community assets" (28%), water and sanitation (22.5%), and within-village roads and bridges (20.4%).[4]

Local governments can affect each category in different ways. For the first two categories, leaders can lobby higher-level bureaucrats and politicians for allocations to their GP, while for the latter two, leader efforts in selection of project and implementation will affect provision. For all three categories, corruption by local government leaders may be relevant in reducing how spending translates into resources on the ground. Both effort and corruption are plausibly affected by polity size: smaller local governments may be easier for elites to capture, but also could have more citizen monitoring that results in less corruption and greater effort. Fundamentally, the net effect of local government size is an empirical question, as it is not clear which of these and other mechanisms will end up being empirically relevant.

3 Data

This paper relies primarily on data from the Indian census and administrative data sets from different government departments. These data are primarily at the village-level, which

[3] For context, this is similar to the per capita expenditure on NREGS at the GP over this period (approximately Rs. 440 per person), for which funding comes from the Ministry of Rural Development rather than the GP budget.

[4] See Table A1 for a full breakdown of all expenditure categories.

is advantageous because villages are nested within GPs, and so outcomes are measured in the same way regardless of which GP the village is associated with. This section gives an overview of each of the data sets, but additional details can be found in the online Appendix.

Census We use data from the Census of India in 1991, 2001, and 2011, and version 1.5 of the Socioeconomic High-resolution Rural-Urban Geographic Data set (SHRUG) to link villages across the waves of the census (Asher et al., 2021). We use population information from the 1991 and 2011 Primary Census Abstract to construct the regression discontinuity running variables and the 2001 data in robustness checks. Data from the Village Amenities component of these census is used to measure the presence of village-level infrastructure (e.g. all-weather roads, schools, and electricity). We also use the Houselisting tables of the 2011 census to measure the percent of households in the village who have a toilet, have houses built of particular materials, or have access to amenities such as electricity or in-house water supply.

Matching villages to gram panchayats We match villages to their corresponding GPs based on two sources of data. To match villages to their GP from 2015 to the present, we use the 2021 version of the Local Government Directory from the Ministry of Panchayati Raj. This records the association of each village with its GP as well as the unique village ID code from the 2011 census of India, so is straightforward to link to census data.

To match villages to their GP for the period 1995-2014, we use data from the Department of Drinking Water and Sanitation, which maintains a yearly panel of the association between villages and GPs between 2009 and 2020. We scraped this from their website in January 2021 and matched it to the census data using a multi-round matching process combining population data and fuzzy name matching, which gives a match rate of over 96% (see appendix OA.1.2 on the authors' websites for details).

Socioeconomic and Caste Census We use village-level data from the Socioeconomic and Caste Census (SECC) to measure educational attainment. The SECC surveyed every household and individual in the country in 2012, and records the number of people in each village at different levels of educational attainment (e.g. no education, some primary, etc.) We scraped this data from the SECC website in August 2022.

Mission Antyodaya For village-level information on the delivery of public services in the 2019-2020 period, we scraped data from the Mission Antyodaya initiative website. These data were collected as part of a central government initiative to have regular and systematic village-level data. They are based primarily on administrative data from government departments; for example, the Department of Food and Civil Supplies reports the number of households in each village who hold below poverty line ration cards. We focus on the delivery of eight services measured in these data, which are discussed in more detail in the Results section.

National Rural Employment Guarantee Scheme We scraped publicly available administrative data on the National Rural Employment Guarantee Scheme (NREGS) from the MGNREGA Public Data Portal in March 2021. The Public Data Portal aggregates data on workfare program delivery at an annual level between 2016-2020. This includes information such as the number of projects completed, demand for work, days of work provided, wages, and expenditure on materials for each gram panchayat.

Financial Data To understand the financial resources available to the panchayats, we use data on GP budgets and spending from the Panchayati Raj accounting system (eGram-Swaraj). This contains information on total spending, transfers from central/state government, and local tax revenues for each financial year from 2013 through the present for every GP in Uttar Pradesh.

Election Data We scraped data from the 2015 and 2020 gram panchayat elections from the Uttar Pradesh State Election Commission in January 2016 and December 2021. For each gram panchayat, we observe the vote share and characteristics of each contesting candidate, including education, age, gender, and caste background. In the 2020 elections, it further contains information on the value of assets held by the candidate as well as any criminal history.[5]

Combining Data Sources The majority of these data sets are at the village level and contain the unique village ID code from the 2011 census, so that we can match perfectly across data sets. In cases where that is not available, we use fuzzy name-matching combined with data on population counts that significantly improve the matches (see online appendix OA.1.2 on the authors' websites for details). Table A2 shows the match rates between the different data sets and the census data that is the backbone of our approach. Match rates

[5]Wealth is divided into immovable (illiquid assets such as land or property) and movable wealth (liquid assets such as vehicles and jewelry). Since there is substantial dispersion and many values of zero, we take the inverse hyperbolic sine when using it. Criminal history is recorded as yes, no, and missing. From examination of the original affadavits, criminal history appears to often be strategically omitted relative to fields like education/wealth. We code no criminal record as a 0, missing as a 1, and confirmed criminal records as a 2 to account for this. See the online appendix for more details.

are typically above 95%, and, most importantly, likelihood of matching is unrelated to the discontinuity that we study for all of the data (columns 3 and 4) so matching issues will not bias our estimates.

4 Empirical Strategy

We begin with a sharp RD design based around population cutoffs, where villages above the cutoffs are more likely to be assigned to a less populous GP. For GP composition between 1995 and 2014, the cutoff is a population of 1000 in the 1991 census, while for GP composition after 2015, the cutoff is a population of 1000 in the 2011 census. The following regression discontinuity specification estimates the impact of delimitation into a smaller GP:

$$Y_{v,g} = \gamma_0 + \gamma_1 1\left\{p_{v,g} \geq c\right\} + \gamma_2 f_1\left(p_{v,g}\right) + \gamma_3 f_2\left(p_{v,g}\right) 1\left\{p_{v,g} \geq c\right\} + \epsilon_{v,g} \tag{1}$$

$Y_{v,g}$ is the outcome of interest in village v of GP g. $p_{v,g}$ is the population of village v in the relevant pre-delimitation census, while c is the population cutoff used for determination of treatment. The coefficient γ_1 captures the treatment effect, while the functions f_1 and f_2 reflect a continuous but potentially non-parametric relationship between the running variable and the outcome. Using the data-driven approach of Calonico et al. (2014) and *rdrobust* command of Calonico et al. (2017), we apply a triangular weighting kernel in distance from the threshold, calculate a MSE-optimal bandwidth h with a linear polynomial estimated within the bandwidth on either side of the cutoff, and calculate heterogeneity-robust standard errors clustered at the GP level g.[6]

Figure 1 tests whether the government follows the delimitation rules, showing the relationship between village population in the census and either (i) the population of that village's GP post-delimitation (panels a and b); or (ii) the likelihood that the village is the only one in its GP (panels c and d). Panels a and c shows the relationship for the 1991 village population, while panels b and d plots the relationship for the 2011 village population. Column 1 of Table 1 provides the corresponding estimates in table form: crossing the treatment threshold leads to an average reduction in GP size of 368 ($t = 13.4$) persons with the 1991 discontinuity, while the reduction is approximately 649 persons for the 2015 discontinuity ($t = 20.89$). The larger value for 2015 reflects the larger average GP size in the later period.[7]

[6] The MSE-optimal bandwidth and polynomial will vary for each outcome Y, following Calonico et al. (2018). We show robustness to alternative kernels, functional forms, and bandwidths in the online appendix on the authors' websites in section OA.3, including forcing the bandwidth to be the same in all specifications.

[7] Online Appendix OA.2 on the authors' websites shows why not all villages with a population over 1000

Crossing the population cut-off of 1000 leads to changes along multiple different dimensions. The village will not only belong to a GP with a smaller population, but also one that is more geographically concentrated and may have a more homogeneous population, among other characteristics. That complicates the interpretation since there are multiple bundled treatments, but is still the treatment effect of interest from a policy perspective: any reduction in the size of local political bodies will necessarily entail changes along all of the other dimensions. We thus focus on estimating the sharp discontinuity rather than instrumenting for any one of these treatments like population size in a fuzzy RD design. Section 6 will then return to this point by testing which of these treatments explains the observed effects.

There are two key identifying assumptions for this approach. The first is that the running variable is not manipulable around the threshold, such as if villages attempted to influence census enumerators. This is quite unlikely given the rigorous Indian census process, and no other papers have ever found evidence of census data manipulation in India (e.g, Asher and Novosad, 2020; Burlig and Preonas, 2021). It is even less plausible here since the delimitation was not foreseeable at the time of the preceding census. Nonetheless, we test for continuity of the density of the village population distribution (Cattaneo et al., 2018), as seen visually in Figure A2. The p-value is 0.44 for the 1991 population distribution and 0.36 for the 2011 distribution, indicating that we cannot reject continuity of the population distribution in either case.

A second assumption is continuity across the regression discontinuity threshold of all village-level covariates that may be related to the outcomes. While this assumption can never be proven, we test this for observable baseline variables. Columns (2)-(7) of Table 1 examine variables from the relevant census round preceding delimitation (panel A for the 1995 delimitation with 1991 census data and panel B for the 2015 delimitation with data from the 2011 census), but find no evidence that any vary discontinuously across the population cut-off (see Figure A3 and Figure A4 for the corresponding figures).

The most likely reason why other variables would vary discontinuously across this population threshold is if there are other government programs that use the same threshold to determine eligibility. While we are unaware of any other programs that used population thresholds of 1000 persons in the 1991 and 2011 censuses, appendix C implements numerous robustness checks showing that this cannot explain our results, which we summarize here. The first addresses the concern that there might be all-India programs that use the same threshold. We show that in other Indian states, the population of the GP that a village is assigned to does not vary discontinuously at a village population of a 1000. We

are split into separate GPs. In most cases, it is because the village's neighboring villages lack population to form an independent, contiguous GP themselves.

then use the outcome data for other Indian states to test for discontinuities in the outcomes around the cutoffs. There is no discontinuity in outcomes in states aside from Uttar Pradesh, inconsistent with a national program that targets the same threshold.

While this rules out national programs, the remaining checks rule out programs specific to Uttar Pradesh that make use of the same cutoff. The second check leverages the fact that at the time of the 2015 delimitation, some villages had populations around the cutoff, but were already the only village in their GP. These villages will be unaffected by the delimitation in 2015 since they are already an independent GP, but *would* be affected by any state or national program using the 1000 person population cutoff for eligibility. We do not observe any discontinuity in the studied outcomes, inconsistent with other state programs targeting that cut-off. Third, we leverage an implementation failure, where approximately 15% of Uttar Pradesh districts did not adhere to the delimitation rules. We show that there is no discontinuity in the outcomes of interest in those districts around the population threshold.

The last check leverages how the delimitation of one village affects other villages. Suppose that between 1995-2014, villages a and b were in a GP with village c, but village c's population is just above the population cutoff in the 2011 census. Village c will be split into a GP of its own. If villages a and b remain together (which is virtually always the case in this situation), this will also reduce the population of their GP. However, villages a and b do not have populations near the 1000-person threshold, so any effects that we observe for villages a and b cannot be the result of another state-specific program using the 1000-person cut-off. Estimates are similar using this approach, reinforcing that the findings are related to polity size.

5 Results

5.1 Village-level infrastructure

We begin by examining educational outcomes in panel A of Table 2 using data from the 2001 and 2011 censuses (see Figure 2 for the corresponding figures). In the short-run, there is significantly more educational infrastructure present in villages to the right of the cutoff, where they are 2.5 percentage points (3.2%) more likely to have a primary school and 1.9 percentage points (6.7%) more likely to have a middle school in 2001. This difference expands in the medium run, where those villages become 4.9 percentage points (10.9%) more likely to have a middle school by 2011, although the effect on primary schools is no longer statistically significant at conventional levels ($p=0.11$).

We supplement this with long-run educational data from the Mission Antyodaya survey.

In 2019, Mission Antyodaya collected village-level data on educational infrastructure such as schools and facilities within the schools (toilets, mid-day meal, etc.), and created an education score for each village, which we re-scale to be on a 0-100 scale. The value of this score is 0.095 standard deviations higher in villages to the right of the threshold (p<0.001), indicating that the effects of polity size persist even 25 years after the original delimitation. A nice feature of our setting is that we observe both the short- and long-run, so could capture dynamics that take longer to develop such as catch-up.

Khanna (2022) shows that additional basic educational infrastructure over a similar time period improves educational attainment and earnings in India, so we would expect the same here. We also test for this directly using village-level data from the 2011-2012 Socioeconomic and Caste Census (SECC). Among other outcomes, SECC measures completed education by every member of the village. Consistent with Khanna (2022), the infrastructure gains significantly increase the number of individuals who have completed at least their primary education. The point estimate is economically large considering that only residents of school-going age over the ten or so years prior to 2012 could be affected by this construction. We take this as evidence that the additional school construction does affect actual educational attainment; effects will plausibly increase further over time as a larger fraction of the village is of an age to attend school when the additional facilities were in place.

Panel B of Table 2 examines three other forms of village-level infrastructure. In both 2001 and 2011, the census measures whether the village has an all-weather approach road made out of gravel/tar. Treated villages initially did have higher provision of all-weather roads (*3.1* percentage points, *p=0.009)* in 2001, but this effect disappears by 2011.[8] This could be due to catch-up from large national road-building programs that occurred after 2001 (Asher and Novosad, 2020). We do not observe effects on electrification in either year, potentially because local government leadership may have less scope for affecting that outcome. However, villages to the right of the discontinuity are more likely to have their own Fair Price Shop (FPS), which provides subsidized grains to poor households.

These results show that smaller GPs provide more infrastructure, and in the case of education, the effect strengthens over time. Given that decisions on the allocation of this type of infrastructure are primarily made at higher levels of government, this at least partially reflects a higher level of effort by GP leaders in lobbying politicians and those officials. It may also reflect reductions in corruption, where infrastructure projects in India, and in Uttar Pradesh in particular, are notorious for high levels of corruption by local governments (Wilkinson, 2006; Lewis-Faupel et al., 2016).[9] Corruption affects both whether infrastructure

[8]Asher and Novosad (2020) study an Indian road construction program based on population cut-offs, but those were based on the 2001 census population rather than 1991.

[9]See Wilkinson (2006) for illustrative anecdotes of the corruption of local politicians in Uttar Pradesh

such as schools are ever built as well as whether the construction is of sufficient quality that it remains functional; a nice feature of our outcome variables is that they measure on-the-ground presence in the long-run, and so capture both lobbying and corruption, even if we cannot disentangle the two.

5.2 Welfare programs

In addition to village-level public goods, local governments play a role in many social welfare programs. These fall under the second categorization from sub-section 2.3: programs that are initiated by individual applications or at the village level, but are funded by higher levels of government. Our first measure is based on implementation of the Indira Awas Yojana (IAY; renamed to "Pradhan Mantri Gramin Awas Yojana" in 2015), a national program that provides the rural poor with cash transfers for upgrading the quality of their homes. Since 1985, over 25 million households have been assisted by this program, and local government leadership plays an important role in spreading awareness and assistance in getting funds. We evaluate IAY performance using data on housing stock from the 2011 census, which records the type of roofing and wall materials in each household. The first two columns of panel C of Table 2 find that households in GPs to the right of the discontinuity are more likely to live in houses made out of brick, and less likely to live in houses constructed from lower quality materials like mud or other organic matter (see Figure 3 for the corresponding figures).

Another program in which GP leadership played an important role was the Total Sanitation Campaign (TSC). This began in 1999 and encouraged construction of toilets with information campaigns and cash transfers to cover 80% of the cost, eventually funding the building of one latrine per 10 people in rural India (Spears and Lamba, 2016). TSC relied heavily on GP leaders as an implementing partner on the ground: a 2013 survey of 700 Uttar Pradesh households who had received toilets through TSC found that 99.4% received support from their GP leaders (Planning Commission of India, 2013).

The 2011 census measures the percent of households in each village who own a latrine, which we use as a measure of successful TSC implementation. Column (4) of panel C of Table 2 finds that latrine ownership jumps by 3.3 percentage points around the discontinuity ($p < 0.001$), reducing the extent of open defecation by roughly the same amount (column 5, $p < 0.001$). Since funding came entirely from the state and national government rather than the GP budget (Ibid.), these results are best interpreted as a measure of effort by GP-level

infrastructure spending. These include a US$150 million project where audits found that a third of funds disappeared and only 44% of the program targets were achieved.

politicians in assisting constituents in accessing national-level programs.[10] Since the toilet and housing measures reflect successfully completed projects, they incorporate any diversion of spending as well as the *quality* of implementation, where shoddy construction will be more likely fall apart by the time of the census.

We also find that 1.2 percentage point more households were connected to closed drain sanitation systems ($p = 0.007$), as opposed to uncovered drains or no drainage system at all. This may be related to the Total Sanitation Campaign or the use of GP budgets to connect the expanded toilet network. At the same time, we do not detect effects on whether a household has piped water (column 6 of panel C), perhaps reflecting a lack of programs dedicated to piped water provision at that time. But given the serious consequences of poor sanitation environments for children's health, the sanitation improvements likely have significant public health benefits (Spears and Lamba, 2016; Geruso and Spears, 2018).

Next, we study the implementation of a series of public benefit programs. We focus on data from Mission Antyodaya, which measures the number of beneficiaries of eight government programs in each village of Uttar Pradesh in 2019. This is based on administrative records from the relevant government departments, where the programs are: (1) Below Poverty Line (BPL) ration cards, which entitle the holders to purchase subsidized grains from government ration shops; (2) publicly provided health insurance through the Pradhan Mantri Jan Arogya Yojana; (3) pensions for the elderly, widows and the disabled under the National Social Assistance Programme; (4) household electricity connections through the Saubhagya scheme (Pradhan Mantri Sahaj Bijli Har Ghar Yojana); (5) receipt of a liquified petroleum gas (LPG) connection through the Pradhan Mantri Ujjwala scheme; (6) receipt of a Rs. 5000 cash transfer for pregnant women and mothers through the Pradhan Mantri Matru Vandana Yojana (PMMVY) scheme; (7) receiving housing subsidies (or being on the waitlist for subsidies) through the Pradhan Mantri Awaas Yojana or state-specific schemes; and (8) having a zero-balance bank account (Pradhan Mantri Jan Dhan initiative).

Since these program data are from the post-2015 period, we run the main RD specification using the 2015 delimitation episode with 2011 census population as the running variable. As this is a large set of outcomes, we combine them into a single index following the method of Kling et al. (2001). Column (1) of panel A of Table 3 finds that villages to the right of the cutoff have a 0.09 standard deviation higher score on this index (see panel (a) of Figure 4 for the corresponding RD plot). The remaining columns of the table do this analysis for each

[10] This and the IAY analysis measure end-user outcomes. Table A3 supplements this analysis with data from the Ministry of Housing and Urban Affairs, measuring the construction of IAY homes in 2014-15, and from the 2011 census, recording whether a village participated in the Total Sanitation Campaign. Consistent with IAY and TSC explaining the results, villages to the right of the discontinuity are more likely to construct IAY houses and participate in TSC.

program individually and show that the effects are perhaps slightly stronger for BPL ration cards, LPG, and housing benefits.

Having a broad set of outcomes is advantageous in providing a more complete picture of the effects, where we can rule out that the improvements in infrastructure come at the expense of welfare programs. But more even importantly, we learn about different channels since the processes through which GP leadership affects delivery of welfare programs differs from the infrastructure outcomes in section 5.1. The welfare program outcomes reflect GP leadership helping constituents make claims on the state (Kruks-Wisner, 2018). This may reflect greater familiarity with members of their community in smaller communities for targeting and information dissemination about program existence. They may also have put more effort in helping them file paperwork and overcome bureaucratic hurdles, as documented in accounts of politician assistance in this process (Auerbach, 2019; Kruks-Wisner, 2018).

5.3 National Rural Employment Guarantee Scheme

Finally, we examine the National Rural Employment Guarantee Scheme (NREGS). We use administrative data over the period from 2016-2020 with outcomes including the number of NREGS projects completed, reported demand for work, days of work provided, and expenditure on labor and materials in each year. Nearly all GPs (99%) have at least one project per year, and approximately 15% of households have at least one member working in NREGS in a given year (1.21 workers per household). Each participant ends up working an average of 30.4 days per year, but only 3% reach the maximum of 100 days per year. We normalize all of the outcomes by dividing by total residents for ease of interpretation (e.g. NREGS wages per resident).[11]

Since each of the NREGS outcomes are highly correlated with each other, we focus on an index created following Kling et al. (2001). In panel B of Table 3 and panel (b) of Figure 4, we find that the index is 0.15 standard deviations higher in villages to the right of the discontinuity. The remaining columns show that improvements are consistent across each of the outcomes.

However, one concern is that the changes in these or other services may differ throughout the distribution of households; for example, gains could be concentrated among advantaged groups if smaller polities are more subject to elite capture. A nice feature of these data are that for three of the variables – number of households possessing job cards to work for NREGS, total person-days worked, and proportion of households who received NREGS

[11]The data is also at the GP-level rather than village-level, so this standardizes the measure. This denominator includes all residents, including those who do not work in NREGS.

work – the data break down the outcome by number of scheduled caste (SC) beneficiaries, so we can measure *who* benefits. Panel A of Table A4 reruns the main RD specification with these outcomes (columns (1)-(3) for SC, columns (4)-(6) for non-SC), and finds similar gains for both SC and non-SC workers. While we cannot measure distributional effects of the earlier welfare programs, these findings point towards gains being shared throughout the community.

These outcomes reflect how GP leaders manage social programs at the village-level. NREGS management is one of the main responsibilities of GP leadership, where they propose projects, hire workers, deal with higher-level administrators for funding, and manage project implementation. As with the earlier outcomes, the NREGS measures reflect leader effort, but may also reflect managerial competence. In section 6, we will dig deeper into what explains these findings – is this simply about it taking less effort for a single leader to manage projects in a smaller polity, or are other factors responsible?

5.4 Estimating heterogeneity by population size

The analysis so far has found consistent evidence of increased public service provision in smaller polities. This approach estimated the treatment effect of allocation into a GP with a population of approximately 1000 (the cut-off value) instead of a GP with a larger counterfactual population.[12] However, the treatment effect may be heterogeneous. More precisely, there may be a treatment effect $\tau(x, y)$ of moving a village from GP with a population of x residents into a GP of y residents, where the effect depends on x and y. For example, it could be that the effect of moving a village from a GP of 6000 people to one of 1000, $\tau(6000, 1000)$, is larger than moving from a GP of 3000 to one of 1000, $\tau(3000, 1000)$; if so, policymakers may want to focus on splitting up larger GPs. Following this notation, the previous sections estimated $\int_x w(x)\tau(x, 1000)dx$, where x refers to the GP population if the village were not split and the weights $w(x)$ reflect the prevalence of this value x. While that is still a policy relevant estimate, a policymaker would certainly prefer to know $\tau(x, y)$ when determining polity boundaries.

We use two estimation strategies to understand $\tau(x, y)$. Our first approach is to estimate $\tau(x, 1000)$ for the values of x that are observable in this context. This is possible in the 2015 delimitation because villages will typically remain with the same GP after 2015 if no village within the pre-2015 GP has a population exceeding 1000. Thus for the 2015 delimitation,

[12]The bottom two panels of Figure 1 helps visualize the counterfactual population distribution – it re-estimates the main RD equation with the dependent variable as a binary variable equal to one if the village's post-delimitation GP is within a particular population range, and then plots the coefficients for different ranges. The figure indicates that the counterfactual GP population is typically between 2000 to 3500.

a village's value of x is the population of its pre-2015 GP. We split the sample into bins of pre-delimitation populations (750-1499, 1500-2249, ..., 6000-6749) and re-estimate the main RD equation within each sub-sample.

Panel C of Figure 4 plots the RD estimates and 95% confidence intervals from each sub-sample for the index of service delivery outcomes, while panel D does this for the index of NREGS outcomes.[13] For both outcomes, the treatment effect $\tau(x, 1000)$ increases with the value of x, i.e. the largest treatment effects are found in villages that otherwise have been in very large GPs. The differences are economically meaningful, where the point estimates are around twice as big for values of $\tau(6000, 1000)$ as compared to $\tau(3000, 1000)$, suggesting that the largest benefits are from splitting up the largest units. We return to this when discussing the mechanisms.

Our second approach estimates $\tau(x, x - 1000)$ for different values of x. For this, we leverage a second source of exogenous variation generated by the delimitation rules. If a village is above the population cutoff in the 2015 delimitation, it is supposed to be split off from its pre-2015 GP. This split also affects the other villages that in its pre-2015 GP, where those villages will typically remain together in a GP that is now 1000 people smaller: for example, for a GP that had a population of 5000 before the 2015 delimitation, the loss of a village that is just above the cutoff will reduce the GP population to approximately 4000 inhabitants.

We can thus use the population of other villages in the same pre-2015 GP as a running variable in a regression discontinuity design. For each gram panchayat in existence prior to the 2015 delimitation, we determine the village whose 2011 census population was closest to 1000. We drop those villages from the sample and use their population as the running variable for the other villages in their pre-2015 GP. Panel A of Figure A8 shows a strong first stage: being in a pre-2015 GP with a village whose population was above the cut-off results in being in a post-2015 GP with an average of 324 fewer people ($se = 49.01$) than if that village's population had been below the cutoff.

Panel B of Table A4 and Figure A8 re-run the analysis from Table 3, but use this RD approach instead. Our power not as good with this approach given the smaller sample size, but the point estimates are remarkably similar to those in Table 2, reinforcing the robustness of those findings. However, our main goal is to understand how the treatment effect varies for different polity size changes. For that, we divide the villages into four sub-samples based on their pre-2015 gram panchayat populations (2000-3000, 3000-4000, 4000-5000, 5000-6000) and estimate the RD within each sub-sample; this corresponds to estimating the treatment

[13]Figure A7 plots these for each of the NREGS outcomes separately. This is not possible with the 1995 delimitation because we don't observe pre-1995 GPs.

effect of moving from a GP of population 2000-3000 to one of 1000-2000, from 3000-4000 to 2000-3000, and so on. Panel F of Figure 4 plots the estimated coefficients when looking at the index of NREGS outcomes. There are positive effects for values of x between 2000-3999, but for higher values of x, the effects are insignificant.[14] This points to a threshold effect, where the key is creating sufficiently small polities rather than positive effects from simply getting smaller, as the next section will discuss.[15]

Discussion Putting these analyses together, the gains from decentralization appear to come from creating sufficiently small local government units rather than simply creating smaller polities. The first analysis of heterogeneous treatment effects showed that the effects were largest when shifting a village out of a larger GP. However, the second analysis complicates the picture, indicating that the gains are not simply from getting smaller, but really about creating sufficiently small local governments; above a certain threshold, allocation into smaller units does not have much of an effect. This threshold appears to be around 1000-2500 persons, which may be sufficiently small that leaders can have a more direct connection with their constituents.

A key outstanding question is what underlying mechanisms drive these findings. One aspect of that question pertains to the bundle of treatments that are necessarily part of reductions in polity size – smaller polities are simultaneously less populous, more geographically compact, and have a higher ratio of leaders to citizens, among other characteristics – but which of these treatments accounts for these results? The other aspect connects to broader models of decentralization: why should we expect polity size to have consequences for service delivery, and what conditions must hold for this to occur? Section 6 explores these points further.

5.5 Robustness checks

This section discusses the robustness of our results to alternative explanations. First, as discussed in the Empirical Strategy section, appendix C implements five different tests of whether the government may use the same 1000-person threshold rule for other programs. We find no evidence for this, indicating that our estimates can be interpreted as the results of GP delimitation.

[14] Figure A9 plots this for each of the NREGS outcomes individually. Panel E of Figure 4 does not detect effects for the service delivery index but the standard errors bars are very wide, perhaps reflecting less power when using the alternative RD approach.

[15] Note also that estimates can be combined to to more fully trace out $\tau(x, y)$: e.g., $\tau(x, x-2000)$ can be approximated by the sum of $\tau(x, x-1000)$ and $\tau(x-1000, x-2000)$ under relatively weak conditions.

Second, we check robustness of the five main tables of the paper to alternative RD approaches: an alternative bandwidth selection method (CER-optimal), fixed bandwidths of 100 and 200, quadratic polynomials, and a uniform kernel. We include the most important tables with a bandwidth of 200 in the paper appendix (Table A9 and Table A10); the remainder are available in appendix OA.3 on the authors' websites. The results are broadly consistent with our main tables.

A third concern is spillovers, and in particular, whether some of the measured benefits for treated villages come at the expense of untreated villages. Even if that were the case, our findings would still be informative directionally, showing that smaller polities have greater public good access. However, the net increase in overall public services would be smaller and policy implications would be more ambiguous, requiring consideration of distributional effects. Spillovers are most plausible onto other villages within the same block, as that is the next highest level of government administration and thus where competition for resources would occur.

Appendix D leverages the delimitation rule to test for negative spillovers, measuring whether each block has a larger fraction of villages just above the threshold of 1000 people as opposed to just below. If there were negative spillovers, an untreated village within a block that has a higher fraction above the cutoff (and thus more GPs and more competition for funds) will have worse public services. There is no evidence of negative spillovers for any of the outcomes. One possible reason is if some of the effects come through reductions in corruption and so do not impose a fiscal cost. Another is that the extent of negative spillovers depends on whether there are hard budget caps *and* whether these caps bind in practice. If either condition does not hold, then more spending in one GP does not necessitate losses for another.[16] Appendix D presents evidence that these conditions hold here, particularly since only a small fraction of villages are delimited into a smaller polity, so the additional spending induced by this would not increase expenditure in any of the studied budget categories by more than 1%. Such a small increase can be accommodated by even limited budgetary slack, but may not hold with more extensive reforms to polity size. Thus even though spillovers do not bias our estimates here, governments considering more intensive polity size reforms should think carefully about what this means in the context of their budgets.

[16]Instead, this would increase government debt and so must be paid down by future taxes; the question of whether that cost exceeds the benefits will be discussed further in the conclusion, but depends on the context and spending type.

6 Mechanisms

This section investigates the mechanisms underlying our results. A better understanding of mechanisms can help in determining other contexts to which these estimates will generalize. These tests are also informative as to the empirical relevance of channels that have been proposed in the political decentralization literature. Finally, delving into the mechanisms also speaks to an interpretational challenge with the discontinuity we examine. Allocating a village to its own GP is a bundle of multiple treatments, and we wish to determine which is actually responsible for the measured effects.

6.1 Financing of public goods

Interpreting our results requires understanding the financing of the outcomes. We particularly want to check whether smaller local governments receive more funding on a per capita basis such that our findings may be a mechanical product of fiscal funding formulas. In that case, even though our estimates would be internally valid for India, they would not generalize to contexts with different fiscal rules or demonstrate general mechanisms.

The outcomes we study are funded from a mix of local government budgets and sources like district or block-level government departments. Section 2.3 noted that fiscal formulas should cause larger and smaller GPs to have nearly equal budgets on a per capita basis, but we confirm this with data from the Panchayati Raj accounting system between 2013 and 2020. Table A5 uses our main RD approach to measure whether polity size affects local government budgets. Panel A examines budget categories on a per capita basis, while panel B looks at the total budget. Consistent with the stated fiscal funding formula, smaller GPs do not have larger budgets on a per capita basis (column 1 of panel A). As a result, they have smaller *total* budgets (panel B), which may limit larger projects. However, this would go in the opposite direction of what we find, and so cannot explain our results.

Even though local taxation is limited, taxes levied by local government may be related to local government size; smaller local governments could be more empowered to impose taxes, but also may have poorer capacity to do so. Applying the same discontinuity, column 2 of Table A5 finds no evidence of a relationship between polity size and per capita local taxation. The relatively limited scope of local taxation is both a strength and limitation of our setting. Results may differ in contexts in which local governments have more substantial taxation powers, but limited local taxation is common in low capacity states, and so our results likely generalize quite broadly.

We next consider whether there is a funding formula causes smaller local governments to be mechanically favored for the set of outcomes that are not funded by the local government budget. Our approach here differs, as we already showed that the smaller local governments receive more spending per-capita for programs like NREGS. However, the question is whether funding rules *mechanically* produce this effect, which would not generalize to other contexts, or the higher spending reflects general aspects of polity size, such as how it can affect local politician effort and performance.

To do this, we consider the types of rules that could mechanically lead to smaller polities receiving more funding on a per capita basis: (1) GPs receive fixed allocations that are invariant to population, so smaller GPs get more per capita; (2) absolute spending caps, which ends up working similarly to fixed budgets ; and (3) spending floors, such that GPs at the floor will spend more per capita if they are smaller. These possibilities can be tested with visualizations of program-specific allocations. We focus on NREGS and the set of eight individually-targeted services since those are our best measures of funding allocations. For NREGS, we measure the total amount spent in a given year. For the individually-targeted services, we approximate spending with the total number of beneficiaries since the per-person spending within a program will be similar (e.g. food entitlement for a ration card is fixed).

Panels A and B of Figure A10 plot the total GP-level spending and number of beneficiaries respectively. There is substantial variation, indicating that the first hypothesis of population invariant budgets is incorrect. Panels C and D test for caps or floors. For each GP, we calculate the maximum amount of spending (beneficiaries) among other GPs within the same block and the GP's spending (beneficiaries) as a fraction of that amount. If there were binding spending caps, then there should be bunching at a value of 100%; instead, there is no more mass there than would expected by chance. Similarly, if there were floors, we would expect bunching at non-zero floor amounts, but instead there are many zeros. Finally, panels E and F plot per capita spending. The dispersion across GPs is meaningful, where per capita NREGS spending amount is five times higher in the top quarter than the bottom. Much of the variation is likely due to the mechanisms through which GP leadership can affect service provision in their area, as the next section discusses.

6.2 Civic engagement and political competition

Some of the most important ways through which the polity size may affect government performance come through politics. One possibility is that smaller polities may have higher levels of citizen engagement and political competition. Mookherjee (2015) notes "[t]he possibility that [political participation and competition] may be affected by decentralization in the long run has been largely ignored", so we address this gap using two rounds of election data (2015 and 2020).

Panel A of Table 4 uses our main RD specification to examine how allocation into smaller GPs affects political outcomes. Column (1) of panel A finds that voter turnout is significantly higher in smaller GPs, with a turnout increase of 1.9 percentage points in villages allocated to smaller GPs.[17] Turnout is likely correlated with other forms of political participation that may motivate leaders, but we cannot observe. Panel D of Table A6 shows that the turnout effects are largest in villages that would otherwise have been in large GPs, which is consistent with this channels being related to the heterogeneous effect estimates from section 5.4.

The next columns examine whether citizens are more likely to stand as candidates for pradhan in small polities, a more personally costly form of civic engagement. Likelihood of running for office increases by around a quarter, with an increase in 1.13 candidates per 1000 over a dependent variable mean of 5.26. Column 3 shows that estimates are similar if we restrict only to candidates who gain at least 5% of the vote share, indicating that this is not just about attracting marginal candidates. Columns 4-6 examine the results of the elections. By some measures, the elections are less competitive, where the average number of candidates declines, a Herfindahl index of vote concentration increases, and the margin of victory is larger. This is likely because there are fewer factions in smaller GPs. It could also be the case that it is easier for voters to coordinate on candidates when they have more information on them, as is likely be the case in smaller jurisdictions.

Putting this all together, there is greater political participation in smaller local governments. This can pressure leaders to perform and likely explains some of our results, although we cannot quantify the magnitude.

6.3 Information and political selection

Another important political channel is selection and discipline of leaders. Seabright (1996) points out that leader actions are more observable in smaller communities, which can make it more expensive to deviate from voter preferences. It may also be easier to enact social sanctions on non-performing leaders or observe candidate quality prior to elections. However, smaller polities have fewer potential candidates, and so those who run may mechanically be

[17]A concern with this approach is that turnout is measured at the GP-level, but we analyze it at a village-level. As a result, the value of the turnout variable for villages to the left of the discontinuity includes turnout in other villages in the (larger) GP. If turnout were systematically lower in those other villages, this would bias the estimates upwards; villages to the right of the discontinuity would have higher GP-level turnout because they are not grouped with other, lower turnout villages. To test this, note this explanation would also predict that splits would cause GP-level turnout to go down in those other villages (as they would no longer be grouped with the high turnout villages near the discontinuity). We test this using our alternative RD approach from section 5.4, which estimates the consequences of GP splits for the other villages which remain in the old GP. The first column of Panel A of Table A6 shows that turnout in the other villages does not decrease after a split.

lower average quality. At the same time, the informational advantages in smaller polities could generate better selection from within a given pool of candidates, and thus incentivize better candidates (Dal Bó and Finan, 2018).

Panels B and C of Table 4 examine the characteristics of the individuals who contest and win the elections as pradhan. We use data from affidavits filed by candidates, as in other studies about on Indian elections (e.g., Fisman et al. (2014); Prakash et al. (2019)). For both election cycles, the affidavits record the education of each candidate, while for the 2020 elections, they also record wealth and criminal record. Neither the average or winning candidates' education is related to GP size (column 1). However, both candidates and winners are less likely to have a criminal record (column 2). We also see no effect on education and a decline in criminality in panel C of Table A6, which uses the alternative RD approach (using the population of another village in the same pre-2015 GP as the running variable) to analyze the same outcomes. These results are highly consistent with decentralization theories in which voters in smaller jurisdictions have an informational advantage in observing hard-to-observe characteristics such as criminality (Boffa et al., 2016). Multiple studies have shown that Uttar Pradesh voters prefer candidates with no criminal record (Banerjee et al., 2014; George et al., 2019), so the high information environment in small polities may deter such candidates from running.

6.4 Elite capture

An argument against decentralization has been that local governments could be more susceptible to capture by local elites (Bardhan and Mookherjee, 2000, 2006). In smaller units, the cost of capture may be lower, such as if clientelistic relationships of landowners determine capture (Anderson et al., 2015); it is more likely that a small core hold a high fraction of land in smaller polities. The idea of smaller governments being more prone to elite capture is a common sentiment globally and can be found in other famous writings such as Federalist Paper No. 10 in the United States. In the case of India, B.R Ambedkar, known the father of the Indian Constitution, opposed the creation of decentralized village government due to concerns about capture by higher caste groups.

Measuring the extent to which polity size affects elite capture is challenging due to the difficulty of measuring elite capture. The affidavit data provide two excellent indicators: the wealth and caste status of village heads (and other candidates). Data on broad caste category (general caste, OBC, SC, ST) of all candidates is available for both elections, while wealth is only measured in 2020. Panel C of Table 4 find that leaders elected in villages allocated to smaller local government units are no wealthier and actually *less* likely to be high caste. A lack of elite capture can help explain the positive consequences of smaller local government for public goods delivery, especially among programs targeted to poor households from which elites would get little benefit.[18]

If there were substantial elite capture, benefits may only accrue to upper caste groups; indeed, B.R. Ambedkar's opposition to village councils was due to concerns about discrimination against scheduled castes. Instead, as discussed in section 5.3, Table A4 shows that gains in NREGS access benefits are similar for both SC and non-SC. These results are perhaps even stronger evidence against elite capture, as they capture the consequences for citizens rather than just the identity of representatives.[19]

6.5 Local leadership

Another mechanism commonly cited in favor of decentralization is reduced social distance between leaders and constituents. This may cause leaders to have better information about the local population for targeting of state interventions (Dal Bó et al., 2021; Balán et al., 2022), personal incentives that are better aligned with constituent preferences (Besley et al., 2004), or greater vulnerability to informal sanctions for poor performance.

In the context that we study, having a population to the right of the discontinuity guarantees that the village has a leader from the village itself. However, if the village were to the left of the discontinuity, the pradhan might be from another village. The pradhan is most likely to come from the most populous village within a GP, as that village will have the votes to dominate elections. Given this, a village that was not previously the most populous in its GP will see a substantial chance in the likelihood of electing local leadership as a result of being to the right of the discontinuity. On the other hand, a village that already had the largest population among villages in its GP will not be as affected in terms of whether it has a local leader.

We use this heterogeneity to test for the effect of local leadership, splitting the sample based on whether a village was the largest village in their GP over the period 1995 to 2014 and estimating the main RD specification for the 2015 delimitation exercise within each

[18]This is in the context of reservation, where around half of elections can only be contested by candidates from low caste backgrounds. However, results on caste/wealth are similar when we restrict to contests without reservation (Panel C, columns 3 and 4 of Table A7); if anything, it looks like election winners may be less wealthy in unreserved seats.

[19]Column 5 of panel C of Table 4 also shows that leaders in the smaller polities are less likely to be female, which could have negative consequences for service delivery, especially for female-focused services (Chattopadhyay and Duflo, 2004; Beaman et al., 2009). However, these costs appear to be at least balanced out by the other positive mechanisms to produce net benefits. We do not see a differentially beneficial effect for males for the one outcome that is disaggregated by gender – person days of work in NREGS (Panel A, column 7 of Table A4) – with an approximately 15% improvement for both men and women.

sample.[20] Columns 1 and 2 of Table 5 examine this heterogeneity with indices of Mission Antyodaya (panel A) and NREGS outcomes (panel B).[21] In both cases, the benefits are concentrated among villages that were not previously the largest village in their GP. By contrast, we cannot reject a null hypothesis of no effect on villages that were previously the largest one in their GP for either outcome. As a further test, Figure A11 splits villages into six bins based on the fraction of the GP population they contained prior to the 2015 delimitation (0-30%, 30-40%, 40-50%, 50-60%, 60-70%, and 70-100%), and then reruns the RD within each bin. Although there is some noise, the drop in effect size begins at 50% – where the transition occurs to a village dominating the GP politically – further consistent with the local leader mechanism.[22]

The largest village results also help in addressing two other interpretations of our main results in section 5. One is that those results reflect a zero-sum competition between villages within a single GP. To fix ideas, suppose that there are two villages in a GP, labeled a and b with populations $p_a + p_b = P$. Suppose also that funding allocations were proportional to population, such that a GP containing both villages receives tP, where t is the transfer per person. If $p_a > p_b$, and so village a dominates politically, then village a may receive more than $\frac{p_a}{p_a+p_b}tP$. This means that village b would benefit from being split into its own GP because it would get tp_b, but village a would *lose* resources and thus have worse outcomes. The data is not consistent with this explanation: instead, service delivery outcomes remain roughly the same in villages that were previously the largest village rather than dropping.

Another alternative explanation is that the benefits reflect a higher ratio of leaders to citizens: when a leader serves a smaller group, it is the same cost to provide a higher per-person level of service. However, this would predict improvements even in villages that were previously the largest village in their GP, which we do not see. The explanation also somewhat clashes with the results of section 5.4, as it would predict that reductions in polity population would always improve service delivery, but we only observe that below certain thresholds.[23] While this could play a smaller role, whether or not the leader is locally based appears to be a better explanation. Although we cannot disentangle if this is due to stronger incentives to help their same-village constituents, better information about local needs, or

[20] This approach is better than using the actual residential location of the leader in the pre-delimitation period, which is an endogenous function of candidate quality.

[21] See Table A8 for results on each component of the index.

[22] This approach is limited in only looking at post-2015 outcomes, but the effects are likely to be as strong or even stronger for the other outcomes. Most of post-2015 outcomes are for low spillover public goods, but Besley et al. (2004) shows that leader proximity has an even stronger effect for higher spillover public goods (e.g. schools).

[23] In a separate paper, we directly test for effects of a higher ratio of politicians using discontinuities in how the number of ward councilors is related to GP population (Narasimhan and Weaver, 2022). We also find no evidence of effects there.

other channels, the key take-away for delimitation policy is creating polities that have local leaders.

6.6 Homogeneity and geographical compactness

The discontinuity we study does not only generate less populous GPs, but ones that are more geographically compact and demographically homogeneous. If one of these treatments mediates our results, that is helpful to know in thinking about design of alternative jurisdictional boundaries; e.g. rather than focusing on population, one could instead focus on geographical compactness. To investigate, we leverage variation across delimitation episodes in the extent to which they make polities more compact or homogeneous. We test whether the treatment effect varies along these dimensions; if it does not, we can conclude that this aspect of reduction in polity size does not explain the results.

Given that individuals tend to live near others from the same identity group, the creation of smaller local government units will tend to imply less diversity in the identity of residents. There may be more cooperation in more homogeneous jurisdictions (Anderson, 2011; Bazzi and Gudgeon, 2021) or leaders may be more likely to be of the same identity group as members of the population and thus more likely to help them (Neggers, 2018; Sharan and Kumar, 2021). To test this hypothesis, we exploit variation in the extent to which the caste composition of a village's local government is affected by delimitation – in some cases, villages with different caste compositions are organized into a panchayat, and so splitting off a village creates a more homogeneous polity; in other cases, the component villages were similar and so caste homogeneity is not affected by fragmentation. For this analysis, we can only use the post-2015 data since it requires knowing the counterfactual caste composition of the GP for treated villages. In 2011, we observe the population in each village that is either scheduled caste, scheduled tribe, or from neither group. We calculate fractionalization (Alesina et al., 2003; Bazzi et al., 2019) by caste in the village as well as in the pre-2015 GP. We then determine whether fractionalization would increase if the village were made into its own local government and test for heterogeneity along the margin.[24]

Table 5 examines the indices of post-2015 Mission Antyodaya service delivery (Panel A) and NREGS outcomes (Panel B). We split villages based on whether delimitation would result in more (column 3) or less fractionalization (column 4), and rerun our main RD design in each sub-sample. We cannot reject equivalence of the estimates across these sub-samples,

[24]It would be ideal to observe more precise caste categories (*jati*) for this calculation. However, even this aggregation should pick up one of the key political divisions in Uttar Pradesh over this period, where one of the most important political parties in this time period had its base among scheduled castes (Bahujan Samaj Party).

indicating that this dimension of heterogeneity is not a strong mediator of the observed effects, though with the caveat that this is not the only measure of homogeneity.

Another consequence of creating smaller local governments is that they cover smaller geographical areas. Greater compactness may increase observability of leader actions, and thus promote greater accountability. As with identity group homogeneity, we exploit heterogeneity in the extent to which the delimitation rule increases compactness. A village is often composed of multiple geographically distinct units known as "hamlets", and so even a single village GP may still have distinct and geographically-dispersed hamlets. If geographical compactness is an important mechanism, then the estimated effects should be largest when a village contains a single hamlet, as geographical compactness is the greatest. Columns 5 and 6 of Table 5 split the sample by whether the village is a single hamlet or multi-hamlet village. We cannot reject the equality of the coefficients, and so conclude that increased geographical compactness of the GP is not responsible for the observed effects.[25] This suggests that observability of leader efforts may not be the driving force behind the estimated effect of local leaders; however, observability could be sufficiently high even in multi-hamlet villages so we view this as suggestive at best.

7 Conclusion

Ever since Plato, thinkers have debated the optimal size of political jurisdictions. In recent years, popular sentiment has moved in favor of smaller polities, but the relationship between polity size and governance outcomes is theoretically unclear. We find that smaller polities indeed provide more public services but that the relationship is non-linear. The increase appears to be induced by a particular set of political mechanisms that are triggered when polities are within a certain size range; in the case of rural Uttar Pradesh, that is around one to two thousand individuals. While the estimates are most applicable to other north Indian states, the political mechanisms likely generalize to other democratic states with similar local government structures. These results offer concrete guidance for governments considering how to construct political jurisdictions.

Our analysis has focused on increases in publicly provided benefits, which often entails higher spending. In the case of rural Uttar Pradesh, where public goods are notoriously under-provided, these investments likely have high returns that may exceed their costs in terms of future taxation. This will not be true in settings where public goods provision is

[25] Since this characteristic is a property of the village rather than the counterfactual local government, we can test this for the full set of outcomes rather than just those measured after the 2015 delimitation. We do this in the online appendix (Table OA19), but can only reject equivalence of the coefficients in one case (electricity).

already high, as a proliferation of smaller polities could lead to over-provision. Yet in either case, our results are still informative in measuring the consequences of polity size reforms: our results demonstrate how and why polity size shapes service delivery, but it is a political decision whether the costs are worth the benefits.

Our data does not permit an evaluation of every possible mechanism through which polity size can affect public service delivery, which would be infeasible in a single paper. We do find that political channels, and in particular having a leader from the local community, explain much of the estimated effects. This suggests some scope conditions for generalizing our results: since most of the mechanisms for which there is evidence center around politics, the estimates may generalize best to other contexts with democratic accountability of local leaders. Future research should consider alternate channels through which polity size may affect public services delivery, such as yardstick competition, density of political representation, and corruption, as well as how polity size interacts with other elements of the environment. It may be that the presence of mechanisms that do not turn out to be relevant in our context, such as elite capture, depend on particular institutional features. Understanding these nuances can help in better design of local government systems and is an exciting area for future research.

References

Alesina, Alberto and Enrico Spolaore, "On the number and size of nations," *The Quarterly Journal of Economics*, 1997, *112* (4), 1027–1056.

_ , Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg, "Fractionalization," *Journal of Economic growth*, 2003, *8* (2), 155–194.

_ , Reza Baqir, and William Easterly, "Public goods and ethnic divisions," *The Quarterly journal of economics*, 1999, *114* (4), 1243–1284.

Alsop, R, Anirudh Krishna, and Disa Sjoblom, "Overview of rural decentralization in India," *World Bank report*, 2000, *3.*

Ambedkar, B.R., "Speech on Village Panchayats Bill in Bombay Legislature," 1932.

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

_ , Patrick Francois, and Ashok Kotwal, "Clientelism in Indian villages," *American Economic Review*, 2015, *105* (6), 1780–1816.

Asher, Sam and Paul Novosad, "Rural Roads and Local Economic Development," *American Economic Review)*, 2020.

_ , Tobias Lunt, Ryu Matsuura, and Paul Novosad, "Development Research at High Geographic Resolution," *World Bank Economic Review*, 2021.

Auerbach, Adam Michael, *Demanding development: The politics of public goods provision in India's urban slums*, Cambridge University Press, 2019.

Balán, Pablo, Augustin Bergeron, Gabriel Tourek, and Jonathan L Weigel, "Local Elites as State Capacity: How City Chiefs Use Local Information to Increase Tax Compliance in the Democratic Republic of the Congo," *American Economic Review*, 2022, *112* (3), 762–97.

Banerjee, Abhijit, Donald P Green, Jeffery McManus, and Rohini Pande, "Are poor voters indifferent to whether elected leaders are criminal or corrupt?," *Political Communication*, 2014, *31* (3), 391–407.

_ , Esther Duflo, Clement Imbert, Santhosh Mathew, and Rohini Pande, "E-governance, accountability, and leakage in public programs: Experimental evidence from a financial management reform in India," *American Economic Journal: Applied Economics*, 2020, *12* (4), 39–72.

Bardhan, Pranab, "Decentralization of governance and development," *Journal of Economic Perspectives*, 2002, *16* (4), 185–205.

_ and Dilip Mookherjee, "Capture and governance at local and national levels," *American economic review*, 2000, *90* (2), 135–139.

_ and _ , "Decentralisation and accountability in infrastructure delivery in developing countries," *The Economic Journal*, 2006, *116* (508), 101–127.

Bazzi, Samuel and Matthew Gudgeon, "The political boundaries of ethnic divisions," *American Economic Journal: Applied Economics*, 2021, *13* (1), 235–66.

_ , Arya Gaduh, Alexander D Rothenberg, and Maisy Wong, "Unity in diversity? How intergroup contact can foster nation building," *American Economic Review*, 2019, *109* (11), 3978–4025.

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

Besley, Timothy and Anne Case, "Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition," *The American Economic Review*, 1995, *85* (1), 25.

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

Boffa, Federico, Amedeo Piolatto, and Giacomo AM Ponzetto, "Political centralization and government accountability," *The Quarterly Journal of Economics*, 2016, *131* (1), 381–422.

Burlig, Fiona and Louis Preonas, "Out of the darkness and into the light? Development effects of rural electrification," 2021.

Calonico, Sebastian, Matias D Cattaneo, and Max H Farrell, "On the effect of bias estimation on coverage accuracy in nonparametric inference," *Journal of the American Statistical Association*, 2018, *113* (522), 767–779.

_ , _ , and Rocio Titiunik, "Robust nonparametric confidence intervals for regression-discontinuity designs," *Econometrica*, 2014, *82* (6), 2295–2326.

_ , _ , Max H Farrell, and Rocio Titiunik, "rdrobust: Software for regression-discontinuity designs," *The Stata Journal*, 2017, *17* (2), 372–404.

Cattaneo, Matias D, Michael Jansson, and Xinwei Ma, "Manipulation testing based on density discontinuity," *The Stata Journal*, 2018, *18* (1), 234–261.

_ , _ , and _ , "Simple local polynomial density estimators," *Journal of the American Statistical Association*, 2020, *115* (531), 1449–1455.

Centre for Policy Research, "Analysis of Fund Flows to Rural Local Bodies - A study for the Fifteenth Finance Commission Centre for Policy Research, New Delhi," 2019.

Chakraborty, Pinaki, Manish Gupta, and Rajesh Thakur, "Overview of State Finance Commission Reports," 2018.

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

Chaudhary, Latika and Lakshmi Iyer, "It Takes A Village? Administrative Devolution and Human Development in India," *Working paper*, 2022.

Cremer, Jacques, Antonio Estache, and Paul Seabright, *The decentralization of public services: lessons from the theory of the firm*, Vol. 1345, World Bank Publications, 1994.

Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora, "Do labor market policies have displacement effects?," *The quarterly journal of economics*, 2013, *128* (2), 531–580.

Dahis, Ricardo and Christiane Szerman, "Development via Administrative Redistricting: Evidence from Brazil," *SSRN Working Paper 3125757*, 2021.

Dal Bó, Ernesto and Frederico Finan, "Progress and perspectives in the study of political selection," 2018.

_ and Martín A Rossi, "Term length and the effort of politicians," *The Review of Economic Studies*, 2011, *78* (4), 1237–1263.

_ , Frederico Finan, Nicholas Y Li, and Laura Schechter, "Information technology and government decentralization: experimental evidence from Paraguay," *Econometrica*, 2021, *89* (2), 677–701.

Fisman, Raymond, Florian Schulz, and Vikrant Vig, "The private returns to public office," *Journal of Political Economy*, 2014, *122* (4), 806–862.

George, Siddharth, Sarika Gupta, and Yusuf Neggers, "Can We Text Criminal Politicians Out of Office? Evidence from a Mobile Experiment in India," Technical Report, Working paper 2019.

Gerring, John, Maxwell Palmer, Jan Teorell, and Dominic Zarecki, "Demography and democracy," *American Political Science Review*, 2015, *109* (3), 574–591.

Geruso, Michael and Dean Spears, "Neighborhood sanitation and infant mortality," *American Economic Journal: Applied Economics*, 2018, *10* (2), 125–62.

Jha, Rajesh, "Social-objectives-laden electoral rules of panchayati raj institutions: a political justice perspective," *JournalNX*, 2021.

Khanna, Gaurav, "Large-scale education reform in general equilibrium: Regression discontinuity evidence from india," *forthcoming at the Journal of Political Economy*, 2022.

Kling, Jeffrey R, Lawrence F Katz, and Jeffrey B Liebman, "Moving to opportunity in Boston: Early results of a randomized mobility experiment," *The quarterly journal of economics*, 2001, *116* (2), 607–654.

Kruks-Wisner, Gabrielle, *Claiming the state: Active citizenship and social welfare in rural India*, Cambridge University Press, 2018.

Lassen, David Dreyer and Søren Serritzlew, "Jurisdiction size and local democracy: Evidence on internal political efficacy from large-scale municipal reform," *American Political Science Review*, 2011, *105* (2), 238–258.

Lewis-Faupel, Sean, Yusuf Neggers, Benjamin A Olken, and Rohini Pande, "Can electronic procurement improve infrastructure provision? Evidence from public works in India and Indonesia," *American Economic Journal: Economic Policy*, 2016, *8* (3), 258–83.

Lipscomb, Molly and Ahmed Mushfiq Mobarak, "Decentralization and pollution spillovers: evidence from the re-drawing of county borders in Brazil," *The Review of Economic Studies*, 2016, *84* (1), 464–502.

Mansuri, Ghazala and Vijayendra Rao, "Localizing development: Does participation work?," 2012.

Martinez-Bravo, Monica, "The role of local officials in new democracies: Evidence from Indonesia," *American Economic Review*, 2014, *104* (4), 1244–87.

_ , "The local political economy effects of school construction in Indonesia," *American Economic Journal: Applied Economics*, 2017, *9* (2), 256–89.

_ , Gerard Padró i Miquel, Nancy Qian, and Yang Yao, *American Economic Review*, 2022, *112* (9), 2921–58.

Mookherjee, Dilip, "Political decentralization," *Annual review of economics*, 2015, *7* (1), 231–249.

Narasimhan, Veda, "Decentralization, Rural Development, and Growth: Evidence from India," *Working paper*, 2021.

_ and Jeffrey Weaver, "Political representation and public goods provision," *Working Paper*, 2022.

Neggers, Yusuf, "Enfranchising your own? Experimental evidence on bureaucrat diversity and election bias in India," *American Economic Review*, 2018, *108* (6), 1288–1321.

NITI Aayog, "Multi-dimensional Poverty Index," *NITI report*, 2021.

Oates, Wallace, *Fiscal Federalism*, New York: Harcourt, 1972.

Olken, Benjamin A, "Direct democracy and local public goods: Evidence from a field experiment in Indonesia," *American political science review*, 2010, *104* (2), 243–267.

Pierskalla, Jan H, "Splitting the difference? The politics of district creation in Indonesia," *Comparative Politics*, 2016, *48* (2), 249–268.

Planning Commission of India, "Evaluation study on total sanitation campaign," *Programme Evaluation Organisation, Government of India*, 2013.

Prakash, Nishith, Marc Rockmore, and Yogesh Uppal, "Do criminally accused politicians affect economic outcomes? Evidence from India," *Journal of Development Economics*, 2019, *141*, 102370.

Rao, Govinda, "Rural fiscal decentralization in India: Problems and reform issues," *Decentralization: Institutions and Politics in Rural India*, 2007, pp. 135–153.

Seabright, Paul, "Accountability and decentralisation in government: An incomplete contracts model," *European economic review*, 1996, *40* (1), 61–89.

Sethi, Geeta et al., *Fiscal decentralization to rural governments in India*, Oxford University Press, 2004.

Sharan, MR and Chinmaya Kumar, "Something To Complain About: How Minority Representatives Overcome Ethnic Differences," *Working Paper*, 2021.

Spears, Dean and Sneha Lamba, "Effects of early-life exposure to sanitation on childhood cognitive skills," *Journal of Human Resources*, 2016, *51* (2), 298–327.

Tiebout, Charles M, "A pure theory of local expenditures," *Journal of political economy*, 1956, *64* (5), 416–424.

Vaishnav, Milan and Neelanjan Sircar, "Core or swing? The role of electoral context in shaping pork barrel," *Typescript. Department of Political Science, Columbia University*, 2012.

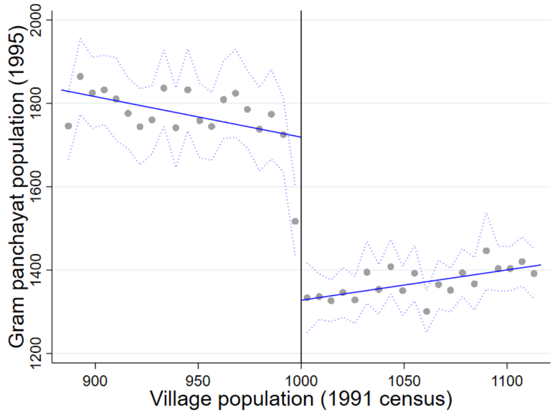
Weaver, Jeffrey, "Political Decentralization and Public Goods Provision: Evidence from India," *Working paper*, 2021.

Wilkinson, Steven I, "The politics of infrastructural spending in India," *Working Paper*, 2006.

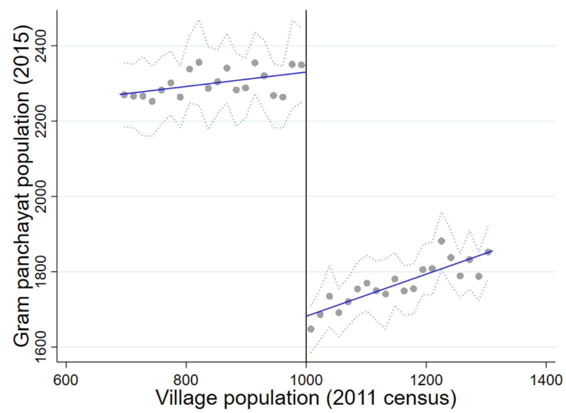
8 Figures

Figure 1: First stage of village population on gram panchayat population in Uttar Pradesh

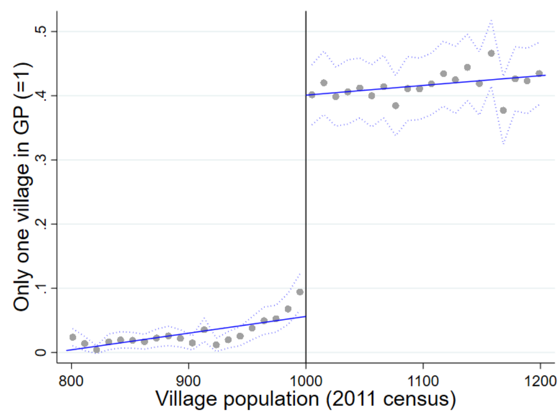
(a) First stage (1991 census)



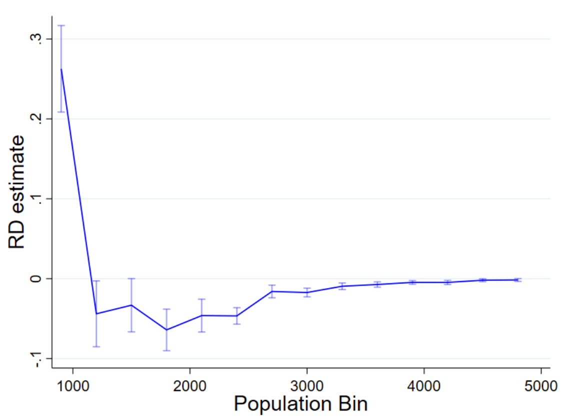
(b) First stage (2011 census)



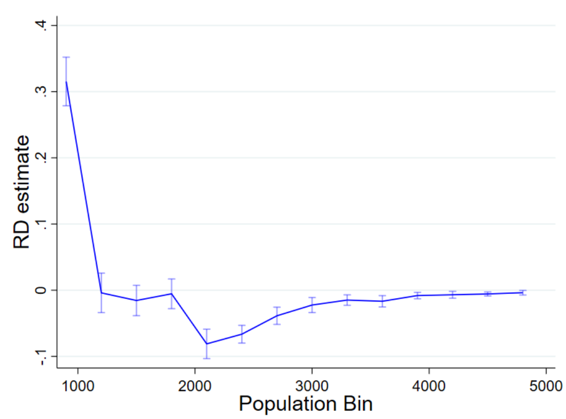
(c) First stage (1991 census)



(d) First stage (2011 census)



(e) GP population (binned, 1991 discontinuity)

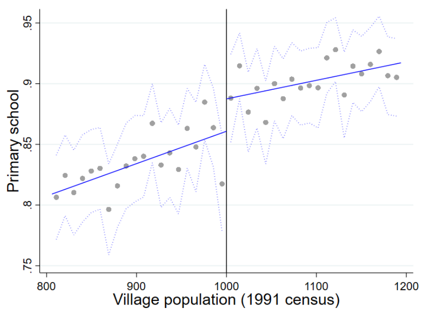


(f) GP population (binned, 2011 discontinuity))

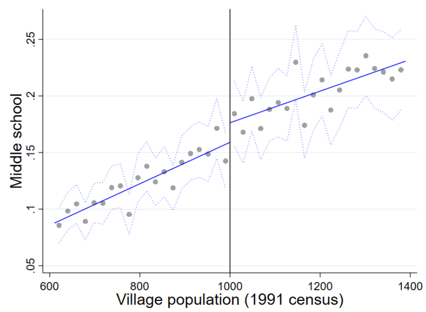
Notes: These figures plot the first stage relationship between GP population and village population. Panels (a), (c), and (e) plot this for the 1995-2015 GP and village population as measured in the 1991 census, while panels (b), (d), and (f) do this for the post-2015 GP and village population from the 2011 census. Panels (a) and (b) show the relationship between total GP population and village population. Panels (c) and (d) show the relationship between whether a village is the only one in its GP and the village population. Panels (e) and (f) rerun the main RD, but where the dependent variable is a dummy variable for whether the GP population that the village is assigned to is in the range (900-1199), (1200-1499),...,(4800-5100). We plot the coefficient and confidence intervals for each of these bins.

Figure 2: Effects on village-level amenities (1991 discontinuity)

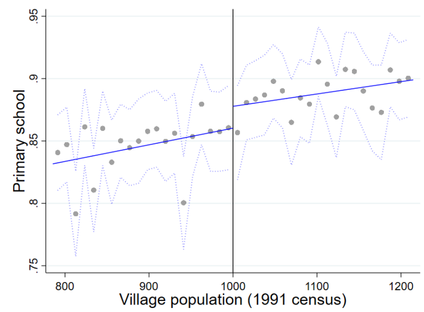
(a) Primary school (2001)

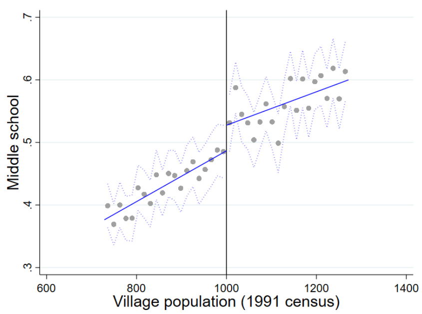


(b) Middle school (2001)



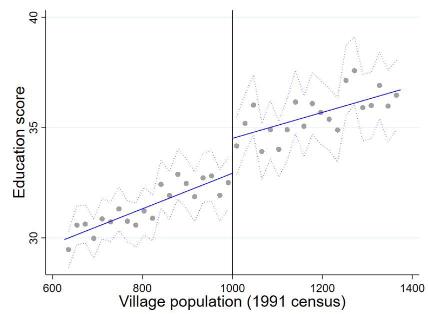
(c) Primary school (2011)



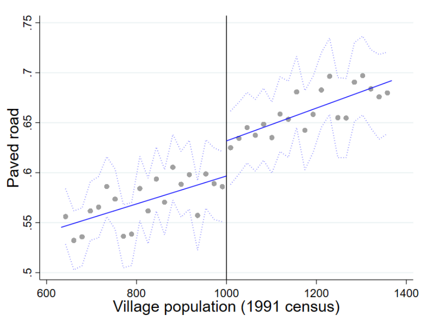


(d) Middle school (2011)

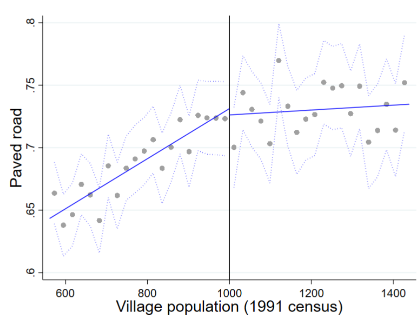
(e) Residents with primary education or higher (2012)



(f) Education score (2019)



(g) Paved road (2001)

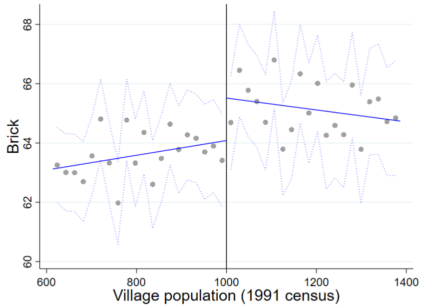


(h) Paved road (2011)

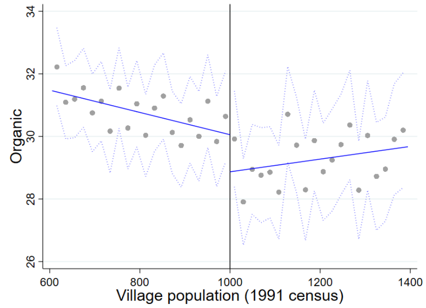
(i) Fair price shop (2011)

Notes: These figures plot the relationship between various characteristics of the village in the 2001 and 2011 censuses and the village's population in the 1991 census. Points to the right of 1000 are above the treatment threshold, while points to the left are below the threshold. Each point represents approximately 480 villages, and the bandwidth is based on the optimal bandwidth selection from Cattaneo et al. (2020). Panels (a) and (b) plot the relationship between the existence of primary and middle schooling (the outcome is a dummy variable taking a value 0 when there is no school, and a value 1 when there is a school) in 2001 against the village population in 1991, while panels (c) and (d) plot the same outcomes in 2011. Panel (e) uses information from the SECC which measures the number of individuals who have a primary education or higher in 2012, while panel (f) uses the Mission Antyodaya data to generate an education measure in 2019. Panels (g) and (h) present the results with having a paved road as the main outcome using a dummy variable taking a value 0 if there exists no paved road, and 1 otherwise, for the year 2001 and 2011. Finally, panel (i) shows the results where the outcome is the existence of a fair price shop in 2011.

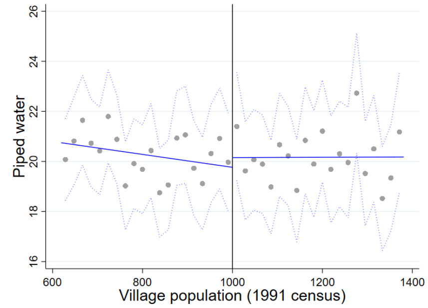
Figure 3: Effects on village-level amenities (1991 discontinuity, continued)



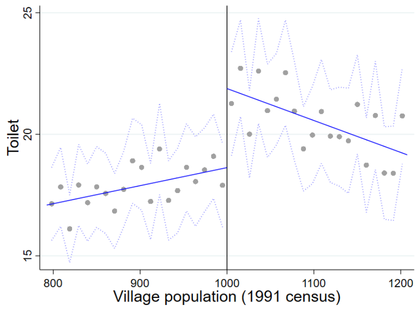
(a) Houses with brick walls (2011)



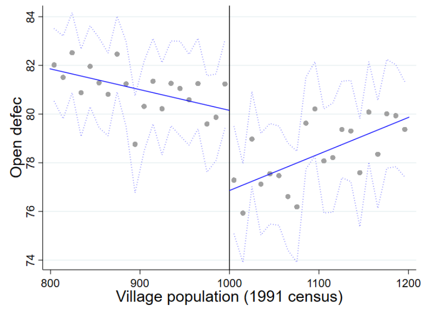
(b) Houses with walls made from organic materials (2011)



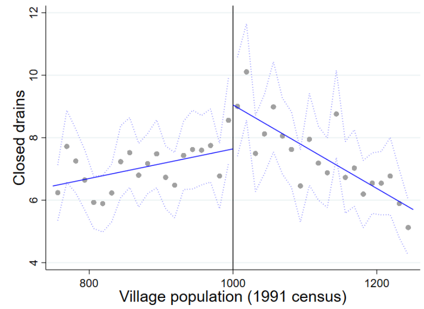
(c) In-house piped water



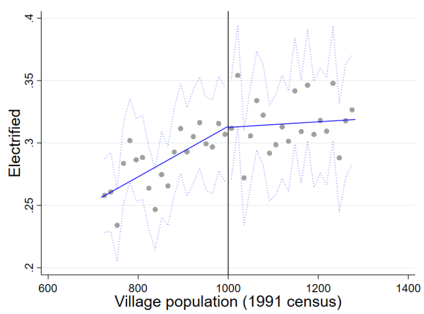
(d) Households with toilets (2011)



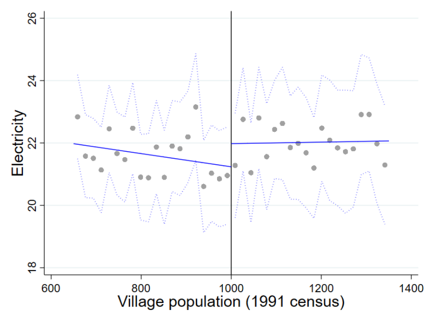
(e) Households who openly defecate



(f) Households with closed drains (2011)



(g) Electrified (2001)

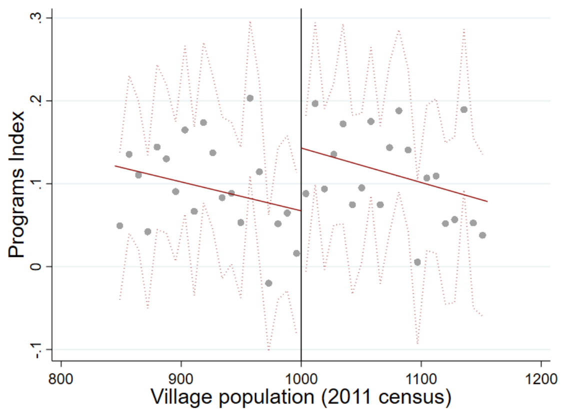


(h) Households with electricity (2011)

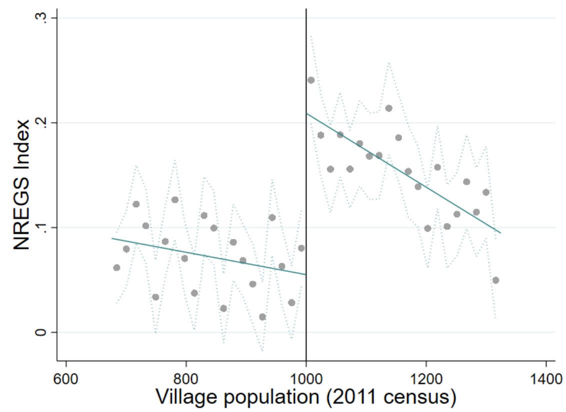
Notes: These figures plot the relationship between various characteristics of the village in the 2001 and 2011 censuses and the village's population in the 1991 census. Points to the right of 1000 are above the treatment threshold, while points to the left are below the threshold. Each point represents approximately 480 villages, and the bandwidth is based on the optimal bandwidth selection from Cattaneo et al. (2020). Panels (a) and (b) plot the relationship between the material used in the construction of house walls in 2011 against the village population in 1991. Panel (a) presents the results for the share of houses built using brick, whilst panel (b) plots the same for houses with organic material (e.g. mud). Panel (c) presents the figure for the outcome of piped water. This is measured by the share of households that have in-house piped water. Panel (d) follows this with a figure presenting the share of households that have a toilet within the village, while panel (e) presents the results for the share of households that openly defecate, and panel (f) shows the share of households with closed drains. Finally, panel (g) and (h) show the outcomes for electricity, with panel (g) measuring share of households that are electrified, while (h) presents the results for the percent of households with electricity.

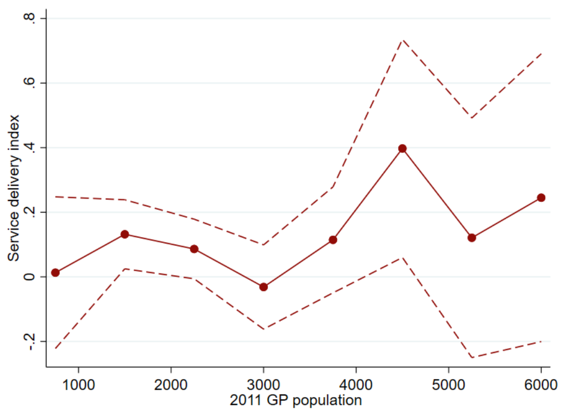
Figure 4: Effects on workfare and individual-level service delivery

(a) Services index

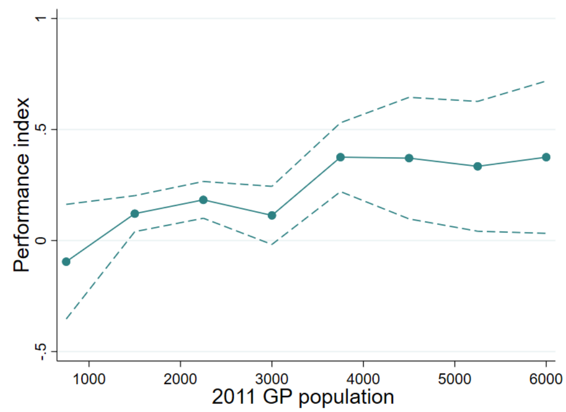


(b) Workfare index

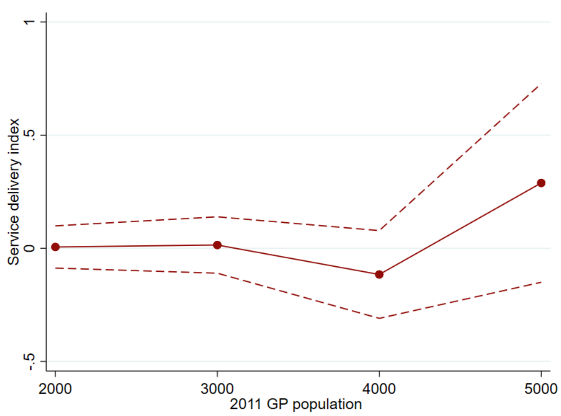




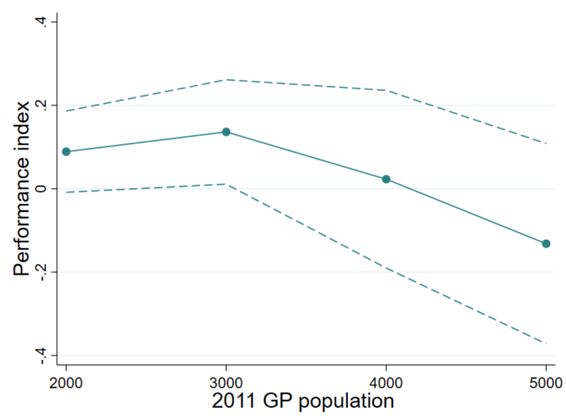
(c) Services index (heterogeneous effects)



(d) Workfare index (heterogeneous effects)



(e) Services index (alternative RD)



(f) Workfare index (alternative RD)

Notes: These figures plot the relationship between outcomes related to services delivery and the village's population in the 2011 census. Panels (a), (c), and (e) examine an index of the delivery of eight services, while the other panels study an index of outcomes related to the National Rural Employment Guarantee Scheme. The first row plots the main regression discontinuity relating village population in the 2011 census to these outcomes. In panels (c) and (d), we divide the sample villages based on the population of their GP prior to the 2015 delimitation (750-1499, 1500-2249, ..., 6000-6749). We run the main RD specification within each of these bins to look at effect of being above the treatment population threshold for villages within the bin. We then plot the coefficients and confidence intervals from each of these regressions in the figure. These estimates can be interpreted as telling us about the treatment effect of moving from that population range into a GP with a population of around 1000. Panels (e) and (f) uses an alternative RD approach based on the population of other villages in the same GP. We divide all of the villages in our sample based on the population of their GP prior to the 2015 (2000-2999, ..., 5000-5999). We run the alternative RD specification within each of these bins: this measures the effect of having another village within your pre-2015 GP be above the the treatment population threshold. We then plot the coefficients and confidence intervals from each of these regression in the figure. These estimates can be interpreted as telling us about the treatment effect of moving from that population range into a GP a population of approximately 1000 less people (e.g. the treatment effect of moving from a GP of 3000 people to one with 2000).

9 Tables

Table 1: First stage and baseline of baseline village characteristics

	Panel A: First stage and balance of characteristics (1991)						
	First stage	Baseline covariates (1991)					
	GP pop.	Primary school	Middle school	Paved road	Electrified	SC pop.	Literate pop.
RD_Estimate	-367.59***	-0.01	0.01	0.02	0.01	-0.01	0.00
	(27.509)	(0.014)	(0.007)	(0.015)	(0.011)	(0.005)	(0.004)
	[0.000]	[0.600]	[0.215]	[0.189]	[0.435]	[0.148]	[0.736]
Dep var mean	1585.886	0.737	0.084	0.420	0.192	0.247	0.277
Bandwidth	252	355	595	493	558	462	323
Effective Obs	9021	18735	31050	23060	29114	24743	14909
	Panel B: First stage and balance of characteristics (2011)						
	First stage	Baseline covariates (2011)					
	GP pop.	Primary school	Middle school	Paved road	Electrified	SC pop.	Literate pop.
RD_Estimate	-649.22***	0.00	-0.00	0.01	0.02	0.00	0.00
	(29.807)	(0.012)	(0.013)	(0.014)	(0.015)	(0.006)	(0.003)
	[0.000]	[0.837]	[0.862]	[0.588]	[0.129]	[0.630]	[0.262]
Dep var mean	2058.658	0.784	0.338	0.663	0.581	0.254	0.564
Bandwidth	468	398	562	435	434	411	355
Effective Obs	25144	20795	26994	23173	22264	21671	16859

This table reports the first stage estimates and balance checks for both delimitation episodes. Column 1 reports the first-stage estimates of the effect of a village having a pre-delimitation census population above the treatment population threshold (1000) on the population of that village's associated post-delimitation GP. The remaining columns check whether any village characteristics in the pre-delimitation census change discontinuously at that population threshold. The first four baseline covariates are whether the village has a primary school, middle school, paved road or electricity , while the remaining two are the fraction of the village's population that is scheduled caste or literate. Panel A run these analyses for the 1995 delimitation and 1991 census, while Panel B does this for the 2015 delimitation and 2011 census. Each specification uses a linear polynomial, triangular kernel, and MSE-optimal bandwidth estimated following Calonico et al. (2017). Standard errors are clustered at the gram panchayat level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Effects on village amenities

Panel A: Educational outcomes						
	2001		2011			2019
	Primary school	Middle school	Primary school	Middle school	Above Primary Ed	Education score
RD_Estimate	0.03**	0.02*	0.02	0.05***	13.54*	1.71***
	(0.012)	(0.009)	(0.012)	(0.015)	(7.984)	(0.434)
	[0.020]	[0.058]	[0.111]	[0.001]	[0.090]	[0.000]
Dep var mean	0.864	0.152	0.864	0.484	668.331	32.900
Bandwidth	298	606	342	451	679	685
Effective Obs	15408	31286	16876	22476	28916	31908

Panel B: Village-level infrastructure					
	2001		2011		
	Paved road	Electrified	Paved road	Electricity	FPS (=1)
RD_Estimate	0.03***	0.01	-0.01	0.84	0.03**
	(0.012)	(0.014)	(0.011)	(0.562)	(0.012)
	[0.009]	[0.681]	[0.532]	[0.137]	[0.031]
Dep var mean	0.600	0.297	0.703	21.748	0.808
Bandwidth	675	425	637	552	352
Effective Obs	36429	22397	33562	29266	18110

Panel C: Household-level infrastructure						
	House			Sanitation		
	Brick	Organic	Piped water	Toilet	Open defec	Closed drains
RD_Estimate	1.31**	-1.13**	-0.01	3.26***	-3.37***	1.25***
	(0.554)	(0.517)	(0.734)	(0.652)	(0.681)	(0.471)
	[0.018]	[0.028]	[0.988]	[0.000]	[0.000]	[0.008]
Dep var mean	64.144	30.223	20.235	18.966	79.902	7.189
Bandwidth	612	628	557	362	366	393
Effective Obs	31916	32596	29266	17194	16969	20104

This table reports regression discontinuity estimates from the main estimating equation of the effect of being above the treatment population threshold on different public-service related outcomes. Panel A presents outcomes related to education: whether the village had a primary or middle school present in 2001 or 2011 (as measured in the census), the number of residents with a primary education or above in 2012 (Socio-economic and Caste Census), and the education score of the village in the 2019 Mission Antyodaya survey. The outcomes in panel B are whether the village has an all-weather road in 2001, whether the village has electricity in 2001, whether the village has an all-weather road in 2011, the proportion of households within the village who had electricity in their home in 2011, and whether the village has a Fair Price Shop within the village in 2011 (all measured in the respective Census round). The outcomes in panel C are the fraction of village residents living in houses made of brick or houses made of organic materials (mud, etc.) in 2011, having in-house piped water in 2011, having a toilet in 2011, primarily defecating in the open in 2011, and with closed drain sanitation systems. See the online Appendix for details of each data set. The running variable is the population of the village in the 1991 census. Each specification uses a linear polynomial, triangular kernel, and MSE-optimal bandwidth estimated following Calonico et al. (2017). Standard errors are clustered at the gram panchayat level and reported below the point estimates. p-values are reported within brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Effect on delivery of services

Panel A: Individual-level programs

	Programs Index	BPL Card	Health insur.	Pension	Saubhagya	LPG	Housing benefits	PMMVY	Jan Dhan
RD_Estimate	0.09***	5.01**	2.20**	-1.34	2.58*	5.63**	5.55**	0.10	4.98**
	(0.033)	(2.412)	(1.057)	(1.075)	(1.401)	(2.450)	(2.187)	(0.134)	(2.048)
	[0.008]	[0.038]	[0.037]	[0.212]	[0.066]	[0.022]	[0.011]	[0.464]	[0.015]
Dep var mean	0.103	109.270	25.841	39.892	50.314	106.895	45.775	2.967	76.798
Bandwidth	275	487	443	435	583	375	368	566	402
Effective Obs	12594	25358	22091	22644	29849	16844	16234	29000	17160

Panel B: Workfare program implementation

	NREGS Index	Work demand	Days worked	Labor expend.	Material expend.	Total projects
RD_Estimate	0.15***	0.01***	0.31***	63.39***	22.67***	0.01***
	(0.024)	(0.001)	(0.050)	(9.448)	(5.564)	(0.001)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Dep var mean	0.111	0.080	2.103	407.795	153.791	0.088
Bandwidth	515	562	520	512	501	499
Effective Obs	119159	131138	121607	119004	107507	117140

This table reports regression discontinuity estimates from the main estimating equation of the effect of being above the treatment population threshold on different welfare program and NREGS outcomes. Panel A presents outcomes related to welfare programs as measured in the Mission Antyodaya data in 2019. Column (1) is an index of the eight programs calculated following Kling et al. (2001). The outcomes are the number of beneficiaries of (1) Below Poverty Line (BPL) ration cards, which entitle the holders to purchase subsidized grains from government ration shops; (2) publicly provided health insurance through the Pradhan Mantri Jan Arogya Yojana; (3) pensions for the elderly, widows and the disabled under the National Social Assistance Programme; (4) household electricity connections through the Saubhagya scheme (Pradhan Mantri Sahaj Bijli Har Ghar Yojana); (5) receipt of a liquified petroleum gas (LPG) connection through the Pradhan Mantri Ujjwala scheme; (6) receipt of a Rs. 5000 cash transfer for pregnant women and mothers through the Pradhan Mantri Matru Vandana Yojana (PMMVY) scheme; (7) receiving housing subsidies (or being on the waitlist for subsidies) through the Pradhan Mantri Awaas Yojana or state-specific schemes; and (8) having a zero-balance bank account (Pradhan Mantri Jan Dhan initiative). In panel B, the first column is an index of the other NREGS outcomes. The remaining columns are the proportion demanding work throuh NREGS, the days worked per person, the total expenditure on wages for labor, the total expenditure on materials, and total number of NREGS projects completed. See the online Appendix for details of each variable. The running variable is the population of the village in the 2011 census. Each specification uses a linear polynomial, triangular kernel, and MSE-optimal bandwidth estimated following Calonico et al. (2017). Standard errors are clustered at the gram panchayat level and reported below the point estimates. p-values are reported within brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Effects on GP election outcomes

Panel A: Political competition						
	Voter turnout	Candidates per 1000	Eff. candidates per 1000	Candidates	Herfindahl index (votes)	Margin of victory
RD_Estimate	1.91***	1.13***	0.88***	-0.90***	0.02***	0.91**
	(0.278)	(0.076)	(0.045)	(0.091)	(0.003)	(0.440)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.038]
Dep var mean	74.773	5.263	3.034	7.534	0.299	11.322
Bandwidth	402	460	439	560	463	530
Effective Obs	37871	43010	39669	48179	42776	46429
Panel B: Candidate characteristics						
	Avg Educ	Avg criminal record	Avg assets (asinh)	General caste perc	Female perc	Avg age
RD_Estimate	0.01	-0.02*	-0.01	-0.01*	-0.02***	-0.12
	(0.025)	(0.010)	(0.072)	(0.008)	(0.008)	(0.125)
	[0.578]	[0.054]	[0.848]	[0.061]	[0.006]	[0.332]
Dep var mean	2.307	0.133	12.572	0.258	0.426	41.026
Bandwidth	561	470	492	533	478	453
Effective Obs	50914	21526	20542	49827	44847	38561
Panel C: Winner characteristics						
	Education	Criminal Record	Total Assets (Asinh)	General Caste	Female	Age
RD_Estimate	0.03	-0.02**	0.00	-0.02**	-0.03***	-0.34
	(0.045)	(0.011)	(0.092)	(0.008)	(0.010)	(0.243)
	[0.547]	[0.042]	[0.986]	[0.027]	[0.002]	[0.168]
Dep var mean	2.518	0.132	12.839	0.282	0.430	41.093
Bandwidth	441	418	518	751	550	472
Effective Obs	40245	17686	22031	69611	48179	43029

This table reports regression discontinuity estimates from the main estimating equation of the effect of being above the treatment population threshold on electoral outcomes in that village's GP. The outcomes in Panel A are related to the overall political competition in the GP, panel B relates to the characteristics of candidates contesting for the position of pradhan, and panel C focused on the characteristics of the candidate elected as pradhan. The data is from the Uttar Pradesh State Election Commission for the 2015 and 2020 elections, including aggregate voting data and candidate affadavits. The outcomes in panel A are voter turnout, number of candidates per 1000 residents of the GP, effective candidates (receiving more than 5 percent of votes) per 1000 residents, the total number of candidates, a Herfindahl index of voting shares, and the winners' margin of victory. The outcomes in Panel B are the average education level, criminal record, age, and asset holdings (inverse hyperbolic sine) of candidates, as well as fraction of candidates who are general caste and female. The outcomes in Panel C are the education level, criminal record, age, asset holdings (inverse hyperbolic sine), caste identity and gender of the person elected pradhan. The running variable is the population of the village in the 2011 census. Each specification uses a linear polynomial, triangular kernel, and MSE-optimal bandwidth estimated following Calonico et al. (2017). Standard errors are clustered at the gram panchayat level and reported below the point estimates. p-values are reported within brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Heterogenous effects based on past GP composition

Panel A: Service delivery performance						
	Largest village		Caste homogeneity		Habitations	
	No	Yes	Decreases /Same	Increases	Single	Multiple
RD_Estimate	0.10***	0.06	0.09**	0.08*	0.07	0.09**
	(0.039)	(0.055)	(0.043)	(0.044)	(0.040)	(0.044)
	[0.008]	[0.256]	[0.027]	[0.080]	[0.102]	[0.034]
Dep var mean	0.087	0.138	0.106	0.111	0.088	0.125
Bandwidth	351	246	340	317	372	320
Effective Obs	9363	4639	7666	7012	8937	7235
Panel B: NREGS performance						
	Largest village		Caste homogeneity		Habitations	
	No	Yes	Decreases /Same	Increases	Single	Multiple
RD_Estimate	0.28***	0.01	0.14***	0.16***	0.13***	0.18***
	(0.032)	(0.048)	(0.037)	(0.033)	(0.030)	(0.033)
	[0.000]	[0.785]	[0.000]	[0.000]	[0.000]	[0.000]
Dep var mean	0.028	0.236	0.131	0.105	0.069	0.162
Bandwidth	510	339	423	587	629	536
Effective Obs	64436	31035	46709	62491	79936	57683

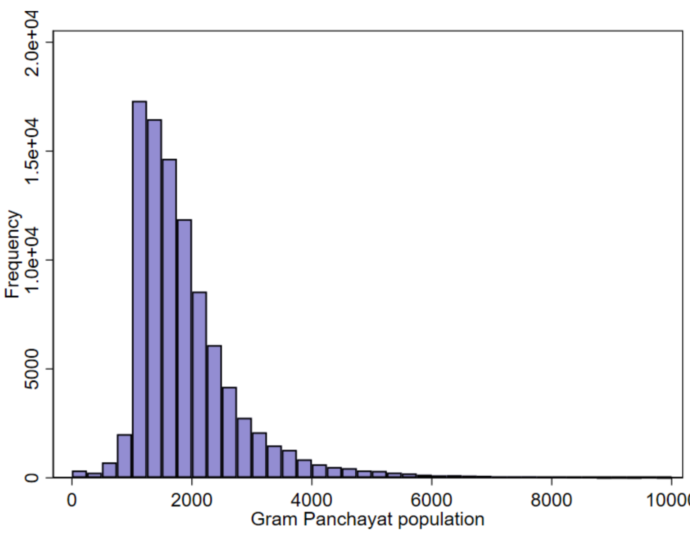
This table reports regression discontinuity estimates from the main estimating equation of the effect of being above the treatment population threshold on different welfare program and NREGS outcomes. The outcome in Panel A is an index of number of beneficiaries for welfare programs as measured in the Mission Antyodaya data in 2019, while Panel B is an index of NREGS program outcomes , both as calculated following Kling et al. (2001). Columns (1) and (2) split the sample based on whether village was the largest village in its GP prior to 2015, and report the RD estimates from each sub-sample. Columns (3) and (4) split the sample based on whether splitting the villages into its own GP would increase or decrease caste fractionalization in its GP. Columns (5) and (6) split the sample based on whether the village has a single habitation or multiple habitations. See the online Appendix for details of each variable. The running variable is the population of the village in the 2011 census. Each specification uses a linear polynomial, triangular kernel, and MSE-optimal bandwidth estimated following Calonico et al. (2017). Standard errors are clustered at the gram panchayat level and reported below the point estimates. p-values are reported within brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix: For Online Publication

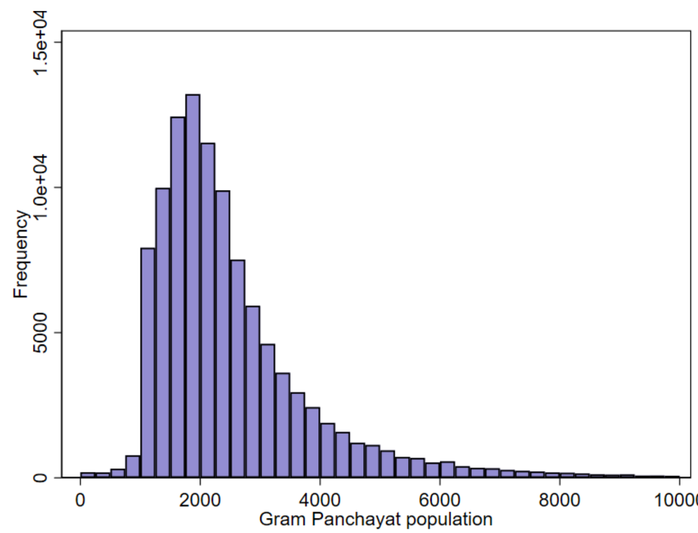
A Appendix Figures

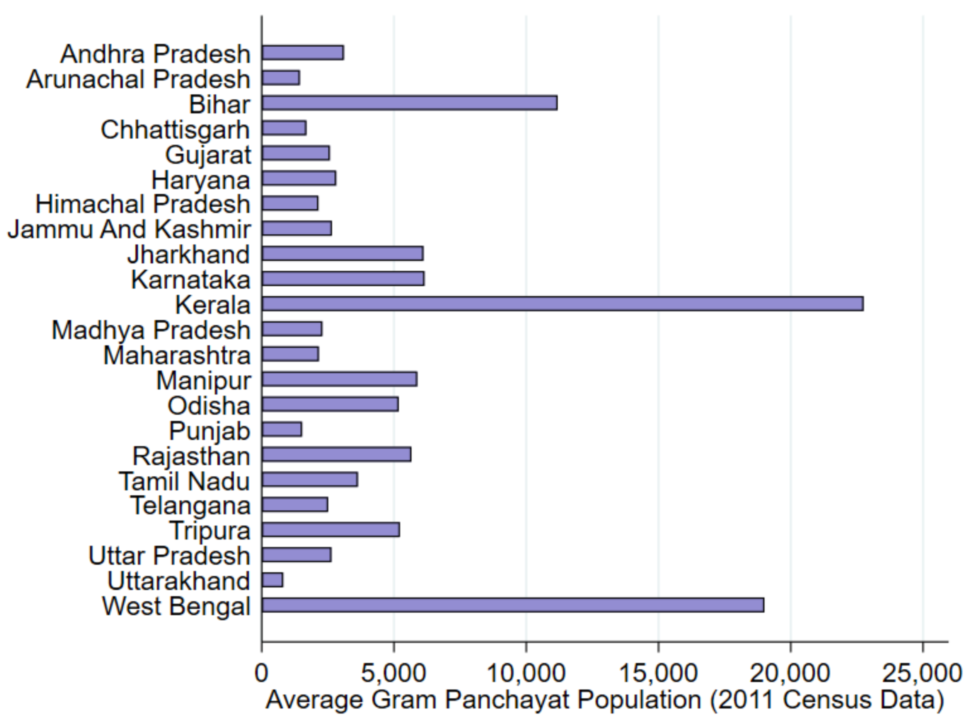
Figure A1: Gram Panchayat population distribution

(a) Uttar Pradesh GP population distribution (1995-2015)



(b) Uttar Pradesh GP population distribution (2015-2021)



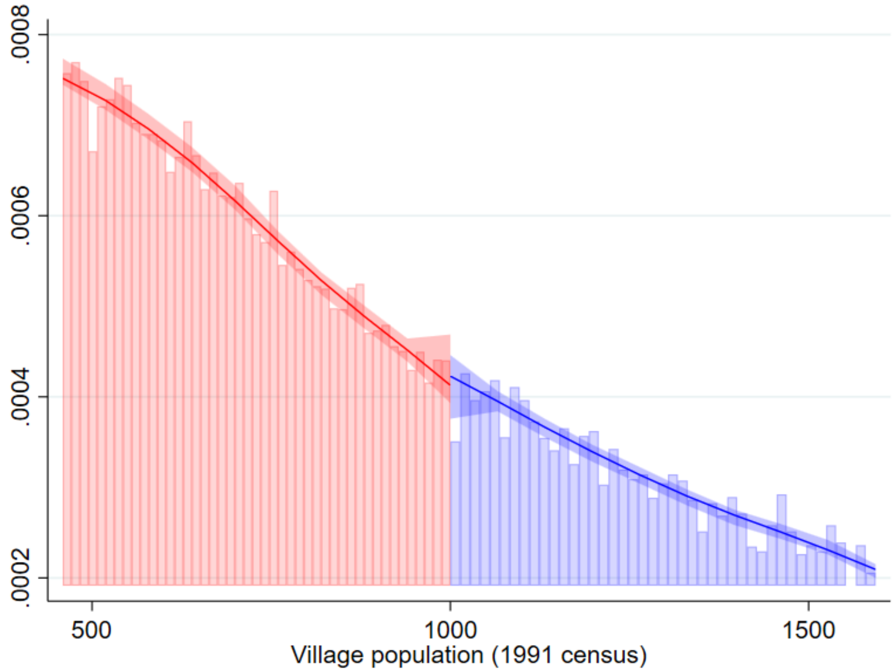


(c) Average size of gram panchayats across states

Notes: Panels (a) and (b) plot the distribution of gram panchayat populations in Uttar Pradesh in 1991 and 2021. Figure (a) uses the 1991 India census data for gram panchayat population for the gram panchayats that existed between 1995 and 2015. Figure (b) uses the 2011 Indian census population data for the gram panchayats that came into existence in 2015. Figure (c) plots the average population of gram panchayats by state. Gram panchayats are defined using the 2021 Local Government Directory, and population is based on the 2011 Indian census (unadjusted for population growth between 2011 and 2021).

Figure A2: Distribution of the running variable

(a) First stage (1991 census)



(b) First stage (2011 census)

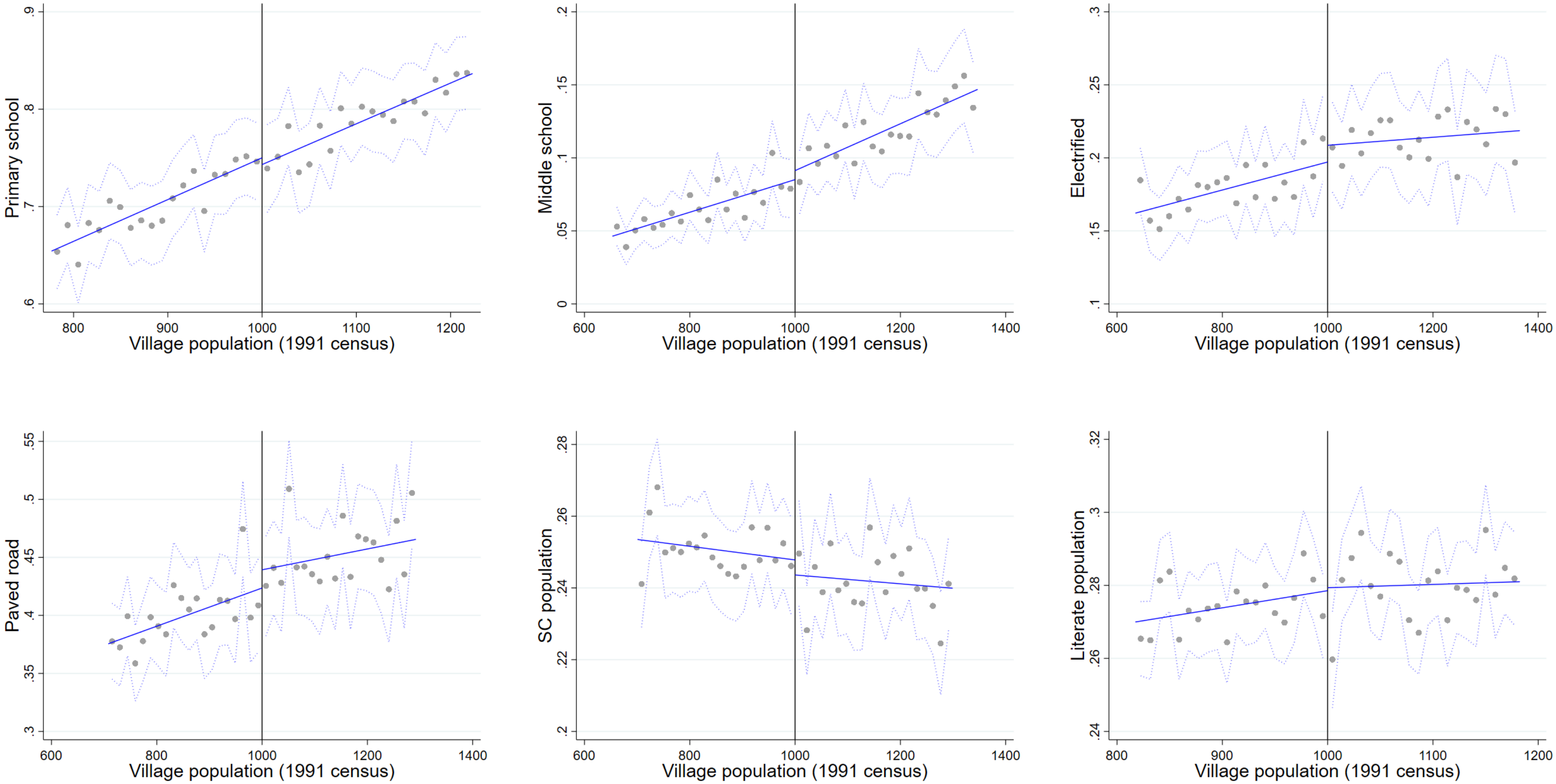
Notes: These figures plot the distribution of village population around the relevant population thresholds. They plot histograms of the population distribution around the relevant threshold and a non-parametric regression for each half of the distribution testing for a discontinuity around the threshold (Cattaneo et al., 2020).

Figure A3: Balance on baseline village characteristics (1991 discontinuity)

(a) Primary school

(b) Middle school

(c) Electricity



(d) All-weather road

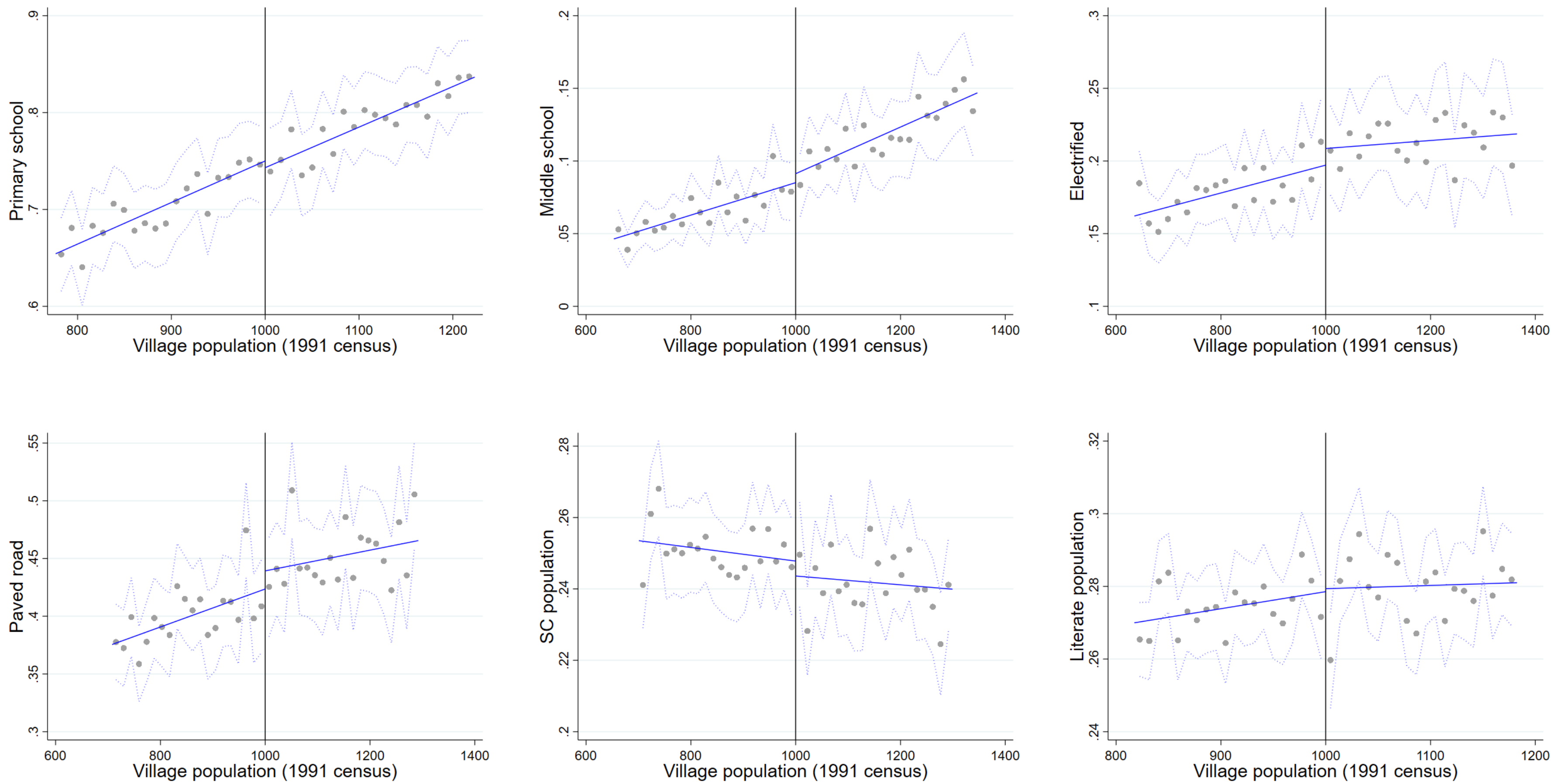
(e) Scheduled caste population

(f) Literate population

Notes: These figures plot the relationship between various characteristics of the village in the 1991 census and the village's population in the 1991 census. Points to the right of 1000 are above the treatment threshold, while points to the left are below the threshold. Each point represents approximately 480 villages, and the bandwidth is based on the optimal bandwidth from the first stage regression.

Figure A4: Balance on baseline village characteristics (2011 discontinuity)

(a) Primary school (b) Middle school (c) Electricity

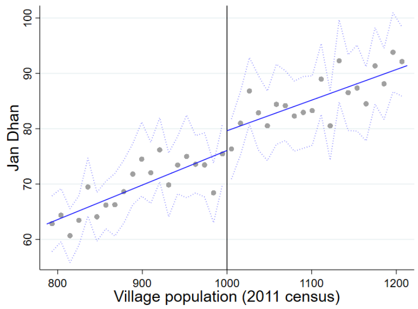


(d) All-weather road (e) Scheduled caste population (f) Literate population

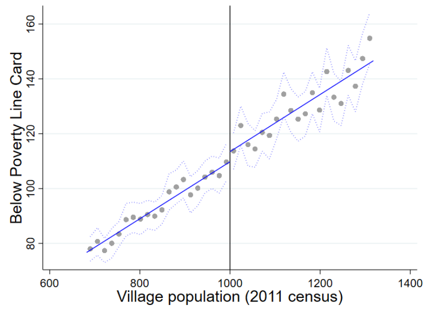
Notes: These figures plot the relationship between various characteristics of the village in the 2011 census and the village's population in the 2011 census. Points to the right of 1000 are above the treatment threshold, while points to the left are below the threshold. Each point represents approximately 480 villages, and the bandwidth is based on the optimal bandwidth from the first stage regression.

Figure A5: Effects on individual-level services (2011 discontinuity)

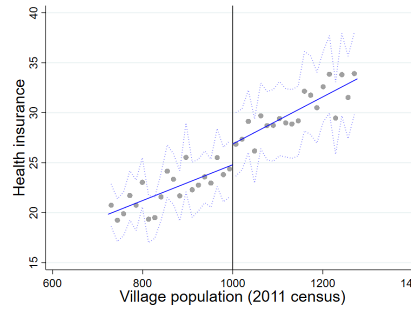
(a) Bank account (Jan Dhan)

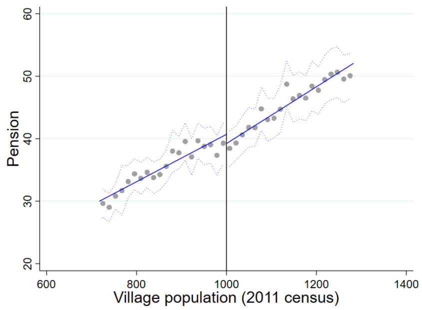


(b) Below poverty line card

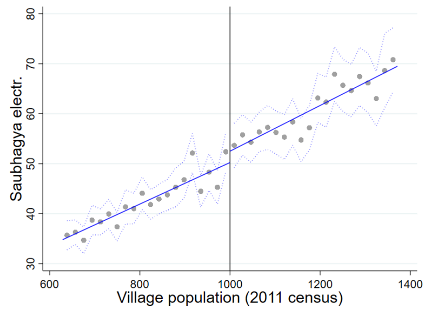


(c) Health insurance

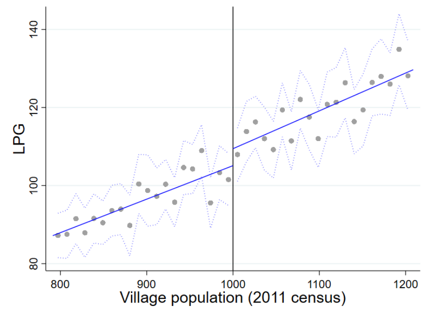




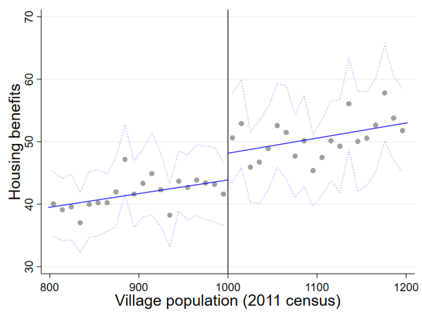
(d) Pension



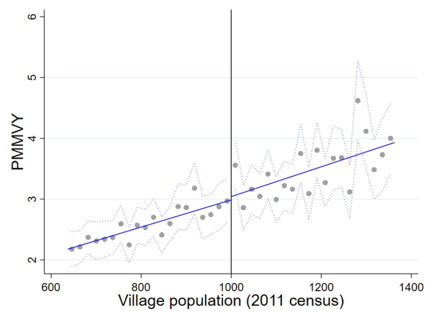
(e) Electricity (Saubhagya)



(f) Clean cooking fuel (LPG)



(g) Housing subsidies

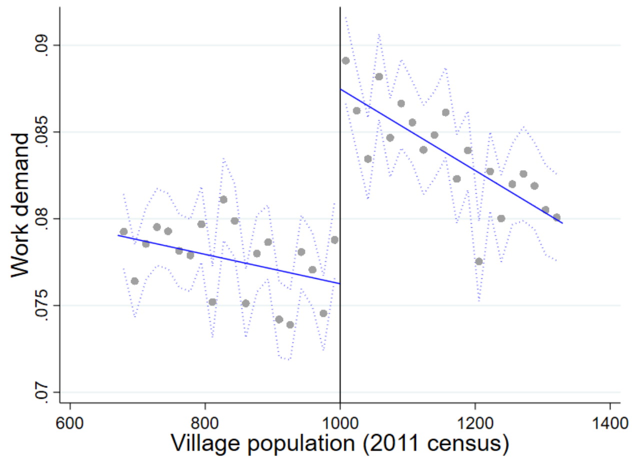


(h) Maternal cash transfer (PMMVY)

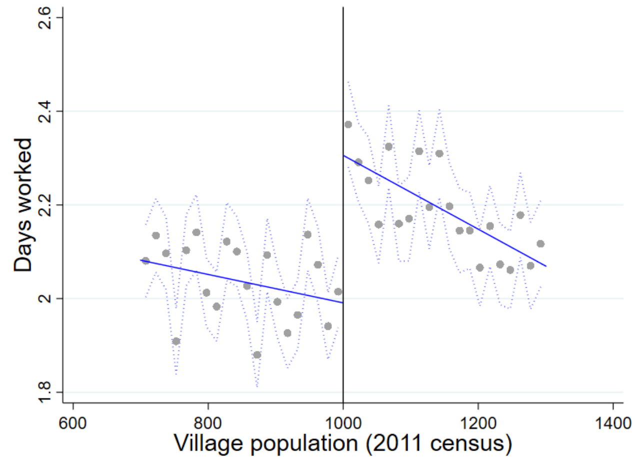
Notes: These figures plot the relationship between various service delivery outcomes in the Mission Antyodaya 2019 data and the village's population in the 2011 census. Points to the right of 1000 are above the treatment threshold, while points to the left are below the threshold. Each point represents approximately 480 villages, and the bandwidth is based on the optimal bandwidth selection from Cattaneo et al. (2020).

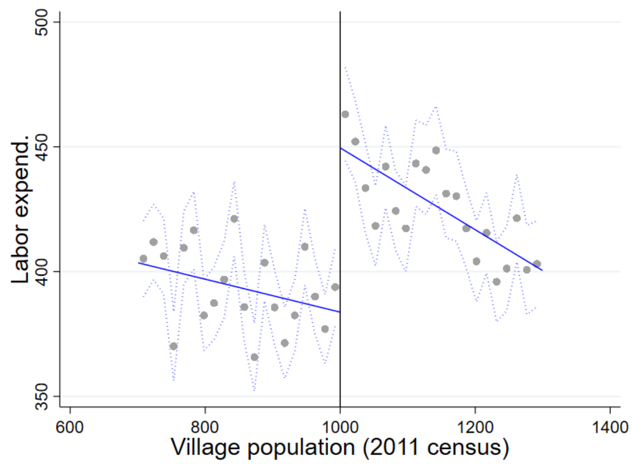
Figure A6: Effects on workfare performance (2011 continuity)

(a) Persons demanding NREGS work

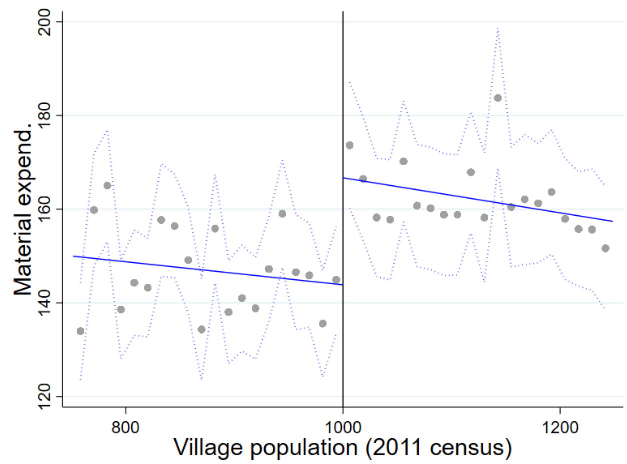


(b) Person-days worked

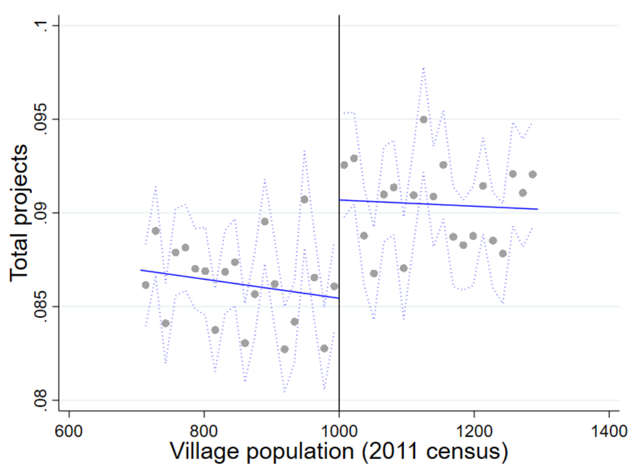




(c) Labor expenditure



(d) Expenditure on materials

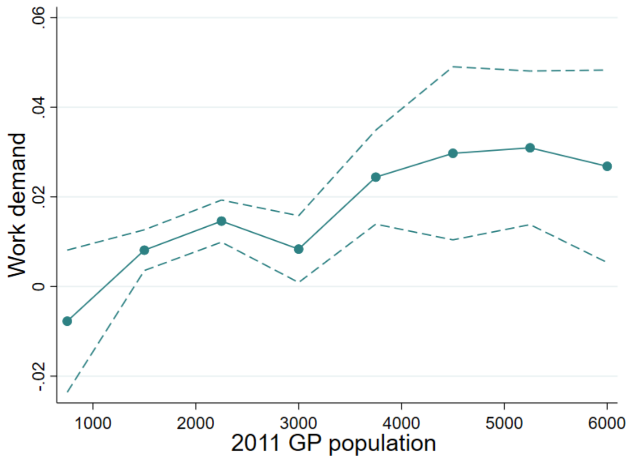


(e) Number of NREGS projects

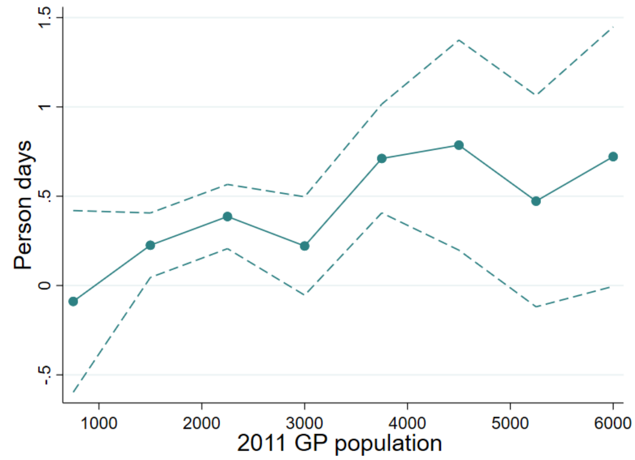
Notes: These figures plot the relationship between outcomes related to the National Rural Employment Guarantee Scheme and the village's population in the 2011 census. Points to the right of 1000 are above the treatment threshold, while points to the left are below the threshold.

Figure A7: Heterogeneity in effects on workfare performance by previous GP population

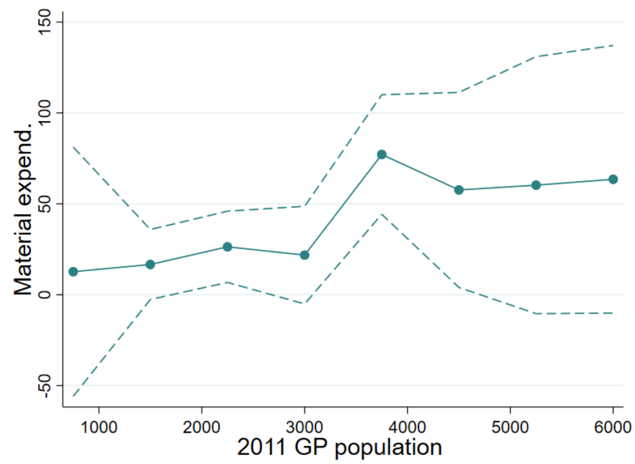
(a) Persons demanding NREGS work



(b) Person-days worked



(c) Labor expenditure



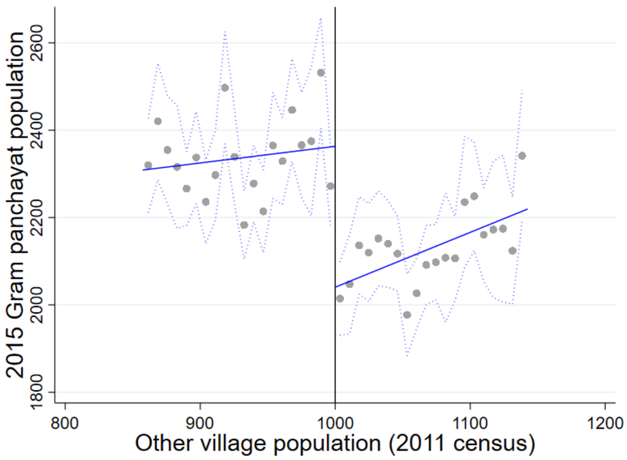
(d) Expenditure on materials

(e) Number of NREGS projects

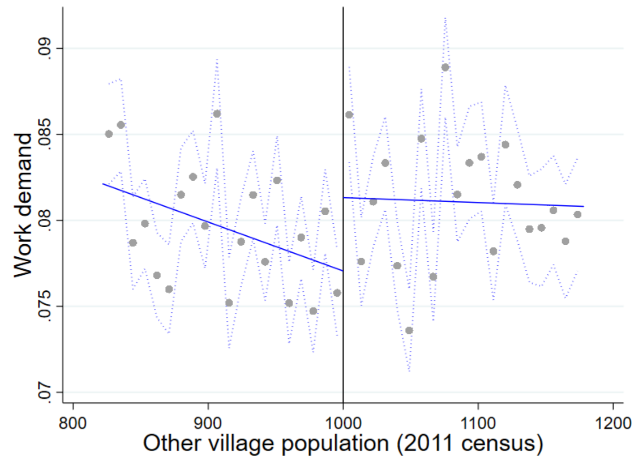
Notes: These figures examine heterogeneous treatment effects on outcomes from the National Rural Employment Guarantee Scheme. For these figures, we divide all of the villages in our sample based on the population of their GP prior to the 2015 delimitation (750-1499, 1500-2249, ..., 6000-6749). We run the main RD specification within each of these bins to look at effect of being above the treatment population threshold for villages within the bin. We then plot the coefficients and confidence intervals from each of these regressions in the figure (e.g., in panel (a), the coefficient for the 750-1499 bin is -0.007, for the 1500-2249 bins, the coefficient is 0.006, etc). Since a village will stay with its pre-2015 GP if its population is below the theshold, these estimates can be interpreted as telling us about the treatment effect of moving from that population range into a GP with a population of around 1000 (the village itself).

Figure A8: Effects on workfare performance (alternative RD)

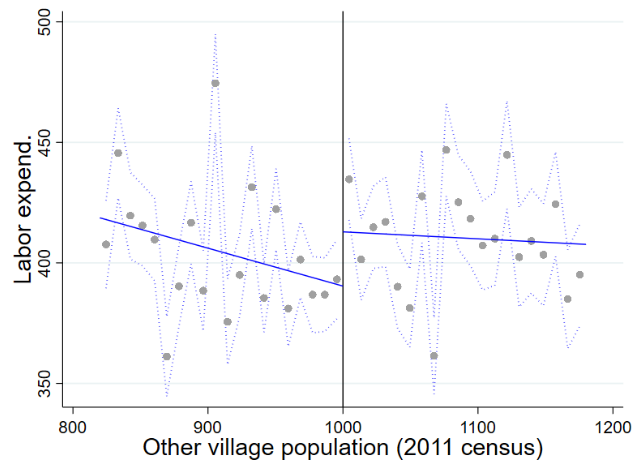
(a) First stage



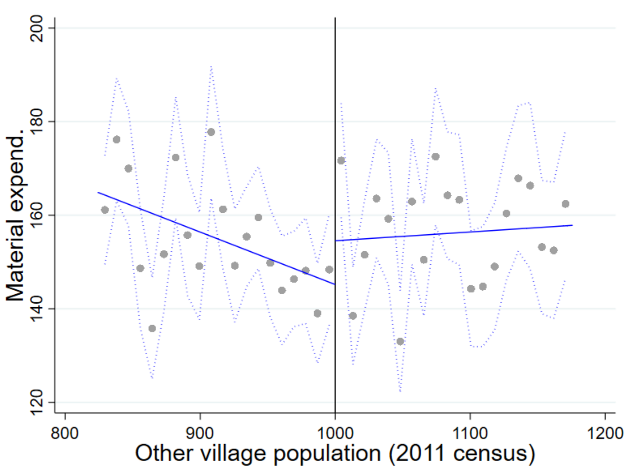
(b) Persons demanding NREGS work



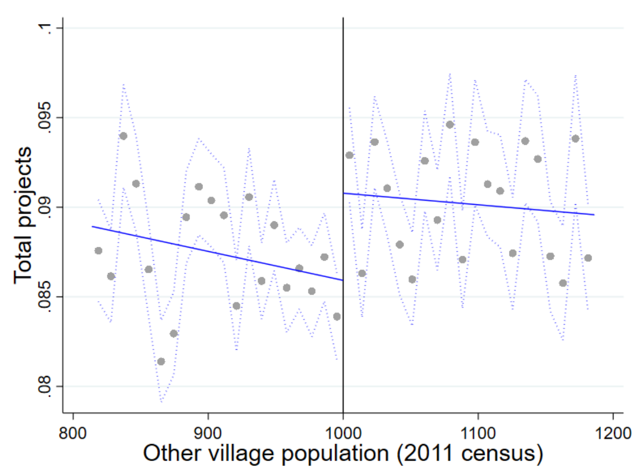
(c) Person-days worked



(d) Labor expenditure



(e) Expenditure on materials



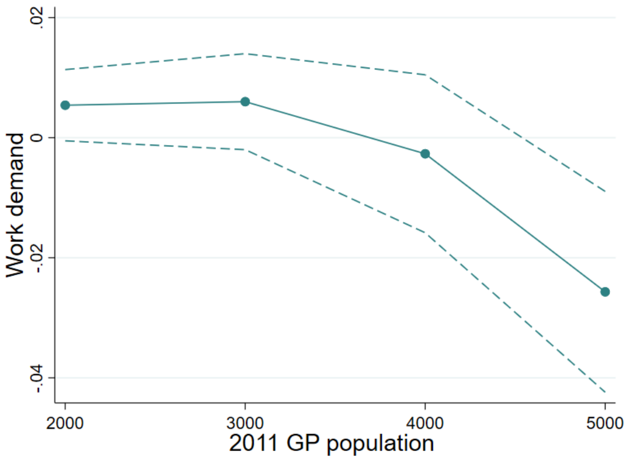
(f) Number of NREGS projects

Notes: This figure uses an alternative regression discontinuity approach to measure the effect of polity size. For each gram panchayat in existence prior to the 2015 delimitation, we determine the village whose 2011 census population was closest to 1000; these are villages who are on the margin of being delimited into smaller GPs in the 2015 delimitation. We drop those villages and then use their 2011 census population as the running variable for the other villages in their pre-2015 GP. If the first village has a population to the right of the discontinuity, this will influence the post-2015 GP population for those other villages because the other villages will almost always stay together in a now smaller GP. This identifies a different local average treatment effect of the effect of post-2015 GP size as both the treatment and population of compliers are different.

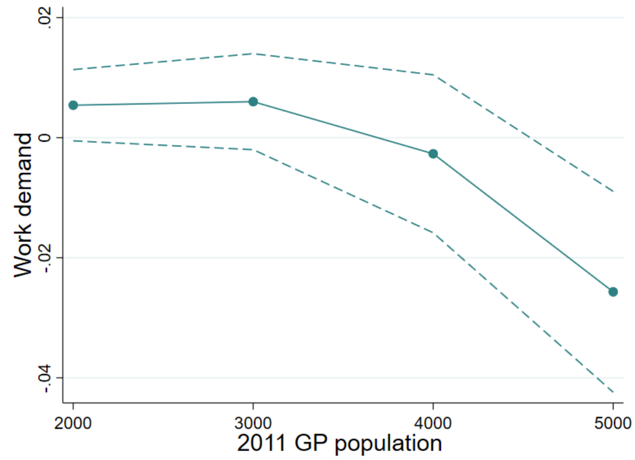
Panel A figure plots the first stage relationship between the population of the omitted village and the 2015 gram panchayat population for the other village of interest. Since the gram panchayat is more likely to be split if the omitted village has more than 1000 people, we observe that those panchayats with a village above the discontinuity have lower gram panchayat populations after delimitation. The remaining panels plot the relationship for different NREGS outcomes.

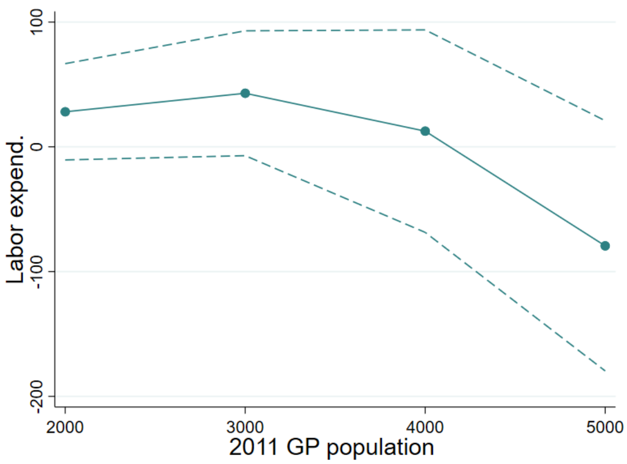
Figure A9: Heterogeneity in effects on workfare performance by previous GP population (alternative RD)

(a) Persons demanding NREGS work

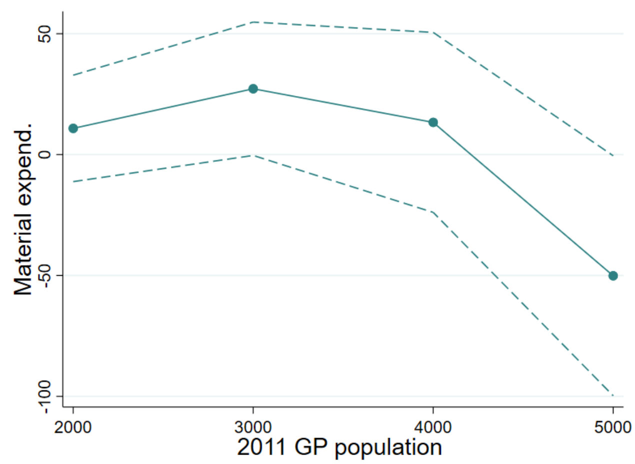


(b) Person-days worked

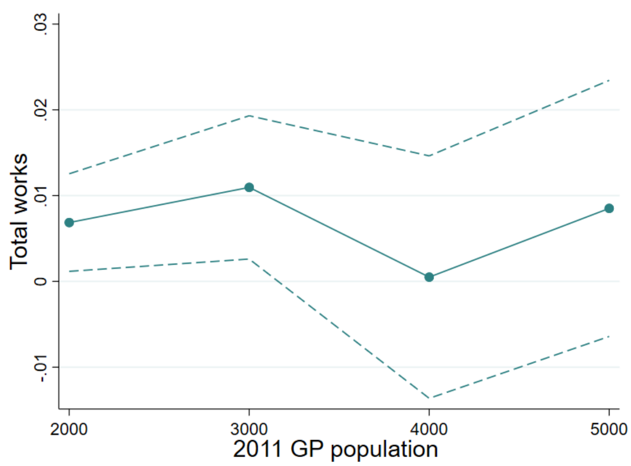




(c) Labor expenditure



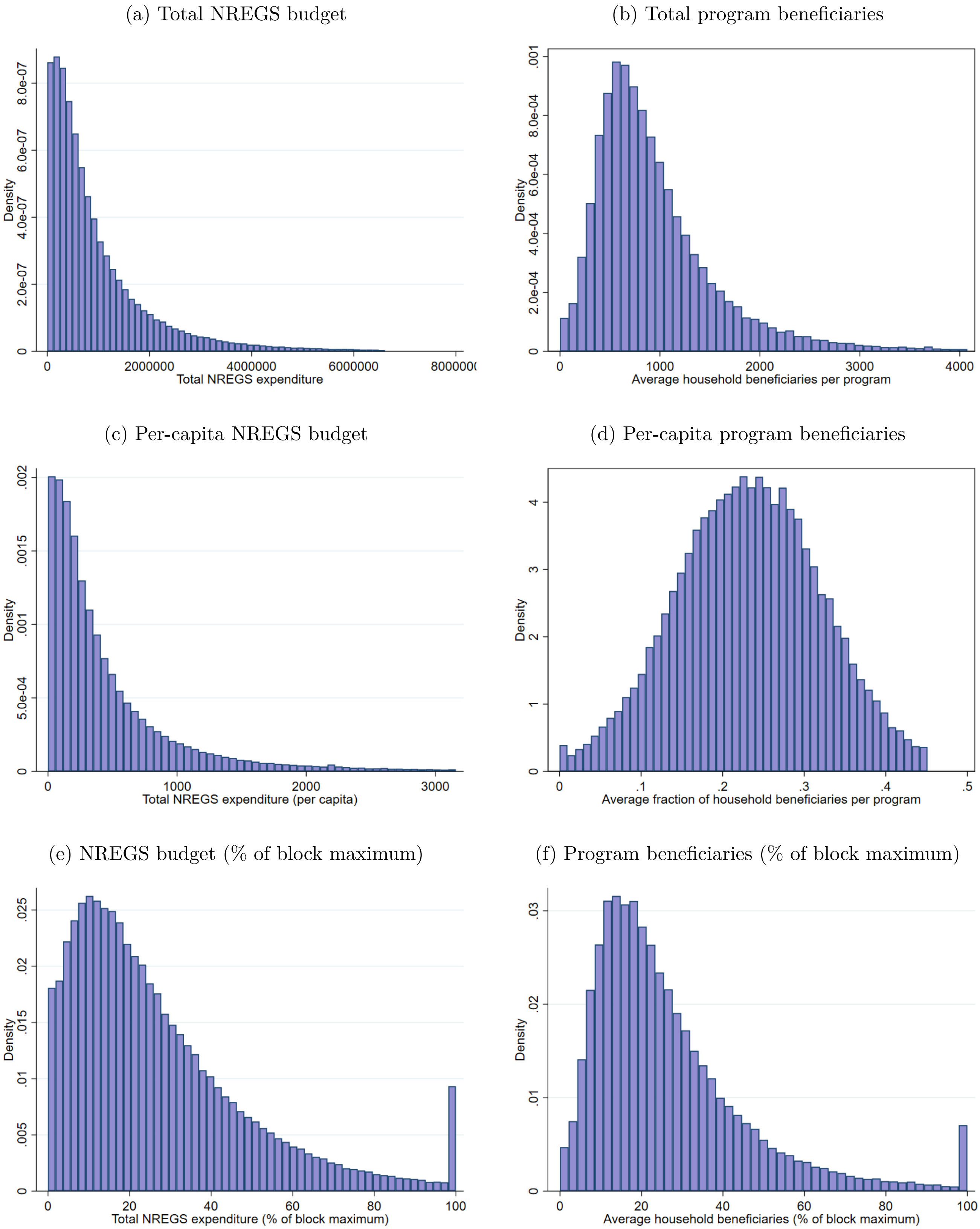
(d) Expenditure on materials



(e) Number of NREGS projects

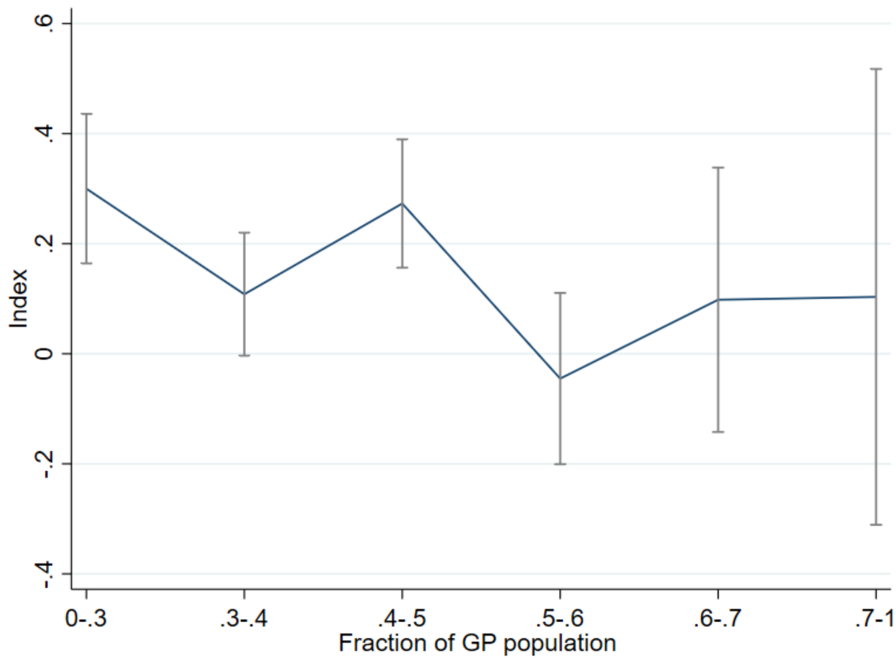
Notes: These figures examine treatment effects on outcomes from the National Rural Employment Guarantee Scheme using our alternative RD approach. For these figures, we divide all of the villages in our sample based on the population of their GP prior to the 2015 delimitation (2000-2999, ..., 5000-5999). We run the alternative RD specification within each of these bins: this measures the effect of having another village within your pre-2015 GP be above the the treatment population threshold. We then plot the coefficients and confidence intervals from each of these regression in the figure. Since the other village will stay with the pre-2015 GP if its population is below the theshold, these estimates can be interpreted as telling us about the treatment effect of moving from that population range into a GP a population of approximately 1000 less people (e.g. the treatment effect of moving from a GP of 3000 people to one with 2000).

Figure A10: Gram Panchayat-level program allocations

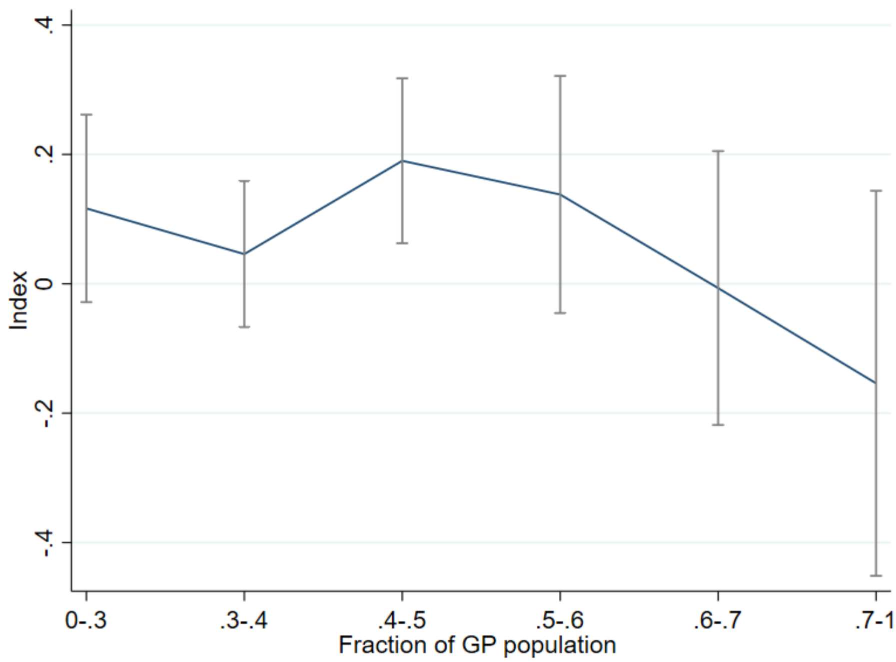


Notes: These figures plot the funding allocations to each gram panchayat in Uttar Pradesh between 2013 and 2020.

Figure A11: GP population share and service delivery



(a) NREGS index



(b) Service delivery index

Notes: For each village, this figure takes the population of the village as a fraction of its total GP population in 2011. It then bins them into six groups: 0-30%, 30-40%, 40-50%, 50-60%, 60-70%, and 70-100% of the total GP population. We then run the 2015 RD specification within each of these six bins and plot the coefficients and 95% confidence intervals in these figures. Panel A plots these six estimated treatment effects for the index of NREGS service delivery, while panel B does this for the index of services measured in the Mission Antyodaya data.

B Appendix Tables

Table A1: The Eleventh Schedule and All-India GP expenditure overview

Panel A: Breakdown of all subjects in the 11^{th} *Schedule*

- Rural housing
- Drinking water
- Fuel and fodder
- Non-conventional energy sources
- Poverty alleviation program
- Libraries
- Public Distribution System
- Maintenance of community assets
- Cultural activities
- Market and fairs
- Fisheries
- Adult and non-formal education
- Welfare of the weaker sections, and in particular, of the Scheduled Castes and the Scheduled Tribes
- Small scale industries, including food processing industries
- Land improvement, implementation of land reforms, land consolidation and soil conservation
- Social, welfare, including welfare of the handicapped
- Roads, culverts, bridges, ferries, waterways and other means of communication
- Rural electrification, including distribution of electricity
- Education, including primary and secondary schools
- Technical training and vocational education
- Women and child development
- Minor irrigation, water management and watershed development
- Animal husbandry, dairying and poultry
- Family welfare
- Social forestry and farm forestry
- Minor forest produce
- Khadi, village and cottage industries
- Health and sanitation, including hospitals, primary health centers and dispensaries
- Agriculture, including agricultural extension

Panel B: Breakdown of average GP expenditure during the 13^{th} *and* 14^{th} *Finance Commissions*

Expenditure Category	13^{th} FC (2013-2015)	14^{th} FC (2015-2018)
1 Building/Community Assets	17.9	24.26
2 Parks, Playgrounds, Burial And Cremation Grounds	4.9	4.14
3 Construction Local Body Roads and Footpaths	27	18.43
4 Maintenance of Local Body Roads and Footpaths	2.3	1.99
5 Water Supply	12.4	11.7
6 Sanitation	10.1	10.74
7 Street Lighting	7.7	10.39
8 Productive Sector	3.1	0.71
9 Welfare	0.9	1.64
10 Salaries, Wages & Employee Pensions	1.9	3.02
11 Office Expenses	2.5	2.6
12 Scheme related expenditure - Central	0.6	0.95
13 Scheme related expenditure - State	0.7	5.29
14 Others	8	4.13

Panel A of this table lists the 29 different sectors that the Central government highlighted as potential areas for devolution for the individual states. State governments selected areas from within this list to delegate to GP responsibility. Panel (b) presents the average expenditure shares (as a %) for GPs for key areas of spending. This information is taken from Centre for Policy Research (2019), a report on fund flows in GPs, which uses a national-level sample of GPs to analyse their expenditure patterns based on publicly available GP finance data.

Table A2: Match rates

	Match rate		Match probability	
	Full sample	In sample	RD estimate	p-value
Panel A: Match to 1991 census data				
DDWS gram panchayat directory	96.3%	97.2%	0.00789 (0.0055)	0.151
Mission Antyodaya	99.9%	99.9%	0.00004 (0.00004)	0.317
Socioeconomic and Caste census	99.8%	99.8%	0.00004 (0.00004)	0.317
Panel B: Match to 2011 census data				
eGramSwaraj	93.9%	96.6%	-0.00409 (0.00635)	0.520
GP elections	86.4%	86.8%	-0.01727 (0.01163)	0.138
Mission Antyodaya	99.9%	99.9%	n/a	n/a
Local government directory	99.1%	99.4%	-0.00122 (0.00198)	0.539
NREGS	87.6%	88.6%	0.00971 (0.01099)	0.377

Columns (1-2) show match rate between the census data and listed data set for all villages and the villages with populations between 800 to 1200. Columns (3-4) test whether the match rate differs around the discontinuity. Column (3) reports regression discontinuity estimates from the main estimating equation of the effect a village of being above the treatment population threshold on whether the village successfully matched to the listed data set: the first number is the coefficient from the RD equation, while the standard errors are reported below that in parentheses. Column (4) reports the p-values from that regression. The running variable is the population of the village in the relevant census round. Note that for the 2011 census-Mission Antyodaya matching, we cannot run the RD specification since the match rate is so high.

Table A3: Additional amenities results: related government programs

	(1) IAY (=1)	(2) TSC (=1)
RD_Estimate	0.039**	0.015*
	(0.02)	(0.01)
	[0.021]	[0.077]
Dep var mean	0.496	0.066
Bandwidth	354	333
Effective Obs	17708	16713

This table reports RD estimates of the effect of delimitation on delivery of two key government programs: IAY and TSC. The running variable is the population of the village in 1991. Standard errors are clustered at the gram panchayat level. The outcome variables are measured as dummy variables. The first is a binary variable for whether a village had any construction under the Indira Awaas Yojana Program in the 2014-2015 fiscal year, as measured by administrative data from the Ministry of Housing and Urban Affairs on IAY construction. Unfortuantely, this is the only year of data available prior to the 2015 delimitation exercise. The second comes from a variable in the 2011 census of India that records whether the village has participated in the Total Sanitation Campaign; however, it is unclear from survey documentation exactly how this was defined.

Table A4: Additional NREGS analysis

Panel A: Effect on delivery of services (by identity group)							
	SC population			Non-SC population			Female
	Job cards	Person-days	HHs worked	Job cards	Person-days	HHs worked	Person-days
RD_Estimate	0.017**	0.391***	0.010***	0.010***	0.346***	0.008***	0.101***
	(0.008)	(0.116)	(0.002)	(0.002)	(0.050)	(0.001)	(0.023)
	[0.030]	[0.001]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Dep var mean	0.214	3.234	0.080	0.130	1.941	0.048	0.722
Bandwidth	438	563	602	607	554	532	496
Effective Obs	100761	122955	133150	137635	129863	124390	116264

Panel B: Workfare using discontinuity in 2011 village population of other villages in the panchayat						
	NREGS Index	Work demand	Days worked	Labor expend.	Material expend.	Total projects
RD_Estimate	0.066*	0.004**	0.109	21.987*	10.035	0.005**
	(0.034)	(0.002)	(0.070)	(13.145)	(7.350)	(0.002)
	[0.050]	[0.047]	[0.121]	[0.094]	[0.172]	[0.012]
Dep var mean	0.120	0.080	2.123	409.057	156.387	0.089
Bandwidth	383	400	374	388	376	410
Effective Obs	96288	99609	95872	97060	97500	101350

This table reports RD estimates of the effect of delimitation on delivery of workfare programs through NREGS. Panel A looks at three NREGS outcomes for which the data breaks down receipt by caste identity of the recipient in columns (1)-(6). In column 7, it examines the one variable disaggregated by gender, which is person-days worked. The running variable is the population of the village in 2011 census. Panel B reruns the analysis from table 3 for NREGS, but using a different empirical approach. For each gram panchayat in existence prior to the 2015 delimitation, we determine the village whose 2011 census population was closest to 1000. We then drop those villages from the sample and use their population as the running variable for the other villages in their pre-2015 GP. This identifies a different local average treatment effect of the effect of post-2015 GP size as both the treatment and population of compliers are different. Standard errors are clustered at the gram panchayat level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Effect of delimitation on GP budgets

Panel A: Per capita budget			
	Total	GP taxation	Additional schemes
RD_Estimate	1.96	0.09	-0.70
	(7.405)	(0.075)	(0.705)
	[0.791]	[0.249]	[0.324]
Dep var mean	579.381	0.242	5.488
Bandwidth	257	147	543
Effective Obs	75772	52656	164856
Panel B: Total Budget (in '000 rupees)			
	Total	GP taxation	Additional schemes
RD_Estimate	-260.32***	-0.04	-4.52***
	(13.614)	(0.132)	(1.300)
	[0.000]	[0.785]	[0.001]
Dep var mean	850.876	0.486	11.374
Bandwidth	403	120	559
Effective Obs	133424	54108	170222

This table reports regression discontinuity estimates from the main estimating equation of the effect a village of being above the treatment population threshold on financial outcomes in their GP. The outcomes in Panel A are measured on a per capita basis, while those in panel B are the total amount for the GP (in thousands of rupees). Column (1) is the total GP budget across all funding categories, Column (2) is the local taxation revenues for the GP. Column (3) is funding from additional government programs. We use the discontinuity based on 1991 population for 2013-2014 and on 2011 population for 2016-2020. We exclude the 2015-16 financial year since that is the year of delimitation. The running variable is the population of the village in the 1991 census for 2013-14 and 2011 census for years after that. Each specification uses a linear polynomial, triangular kernel, and MSE-optimal bandwidth estimated following Calonico et al. (2017). Standard errors are clustered at the gram panchayat level and reported below the point estimates. p-values are reported within brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Effects on GP election outcomes (alternative RD)

Panel A: Political competition

	Voter turnout	Candidates per 1000	Eff. candidates per 1000	Candidates	Herfindahl index (votes)	Margin of victory
RD_Estimate	0.22	0.39***	0.35***	-0.53***	0.01	-0.55
	(0.359)	(0.090)	(0.054)	(0.144)	(0.004)	(0.608)
	[0.545]	[0.000]	[0.000]	[0.000]	[0.113]	[0.370]
Dep var mean	73.439	4.909	2.805	7.677	0.296	11.216
Bandwidth	315	318	267	278	440	400
Effective Obs	37375	36198	32256	32398	46169	38232

Panel B: Candidate characteristics

	Avg Educ	Avg criminal record	Avg assets (asinh)	High caste perc	Female perc	Avg age
RD_Estimate	0.02	-0.03*	0.07	0.01	-0.02*	0.05
	(0.037)	(0.013)	(0.089)	(0.010)	(0.010)	(0.143)
	[0.527]	[0.053]	[0.455]	[0.547]	[0.073]	[0.743]
Dep var mean	2.313	0.134	12.508	0.261	0.430	41.034
Bandwidth	306	315	354	412	434	414
Effective Obs	36287	17196	20003	43848	45175	42374

Panel C: Winner characteristics

	Education	Criminal Record	Total Assets (Asinh)	General Caste	Female	Age
RD_Estimate	0.04	-0.03**	0.07	0.00	-0.02*	0.05
	(0.063)	(0.014)	(0.111)	(0.012)	(0.012)	(0.301)
	[0.571]	[0.043]	[0.530]	[0.952]	[0.072]	[0.879]
Dep var mean	2.523	0.132	12.770	0.283	0.435	41.116
Bandwidth	306	343	368	397	416	411
Effective Obs	34537	17647	19842	43848	44752	43957

Panel D: Voter turnout heterogeneity by counterfactual population

	0-2000	2000-3000	3000-4000	4000-5000	5000-
RD_Estimate	0.03	1.91***	3.11***	5.07***	7.49***
	(0.442)	(0.463)	(0.685)	(1.280)	(1.555)
	[0.954]	[0.000]	[0.000]	[0.000]	[0.000]
Dep var mean	75.544	74.934	74.047	73.561	71.704
Bandwidth	513	357	473	420	491
Effective Obs	14791	13276	6609	1785	1277

This table reports regression discontinuity estimates from the main estimating equation of the effect of being above the treatment population threshold on electoral outcomes in that village's GP. Panels A, B, and C replicate table 4, but where the running variable is the population of another village in the same pre-2015 GP that is closest to 1000 people (2011 census). Panel D divides villages based on the population of their pre-2015 GP (0-2000, 2000-3000, etc). It runs the main RD approach in each subsample with the outcome as voter turnout, where the running variable is the population of that village in the 2011 census. Standard errors are clustered at the gram panchayat level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Effects on GP election outcomes (unreserved constituencies)

Panel A: Political competition						
	Voter turnout	Candidates per 1000	Eff. candidates per 1000	Candidates	Herfindahl index (votes)	Margin of victory
RD_Estimate	1.80***	1.05***	0.85***	-0.97***	0.02***	0.37
	(0.305)	(0.106)	(0.051)	(0.120)	(0.005)	(0.667)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.579]
Dep var mean	74.752	5.264	3.018	7.533	0.299	11.363
Bandwidth	542	448	633	637	394	455
Effective Obs	17397	15635	19034	18540	12923	14587
Panel B: Candidate characteristics						
	Avg Educ	Avg criminal record	Avg assets (asinh)	High caste perc	Female perc	Avg age
RD_Estimate	-0.00	0.00	-0.10	0.00	-0.01	-0.20
	(0.040)	(0.016)	(0.116)	(0.015)	(0.006)	(0.154)
	[0.947]	[0.780]	[0.389]	[0.955]	[0.314]	[0.189]
Dep var mean	2.305	0.134	12.575	0.258	0.426	41.044
Bandwidth	490	451	425	426	767	631
Effective Obs	16813	7621	7310	13242	24071	22028
Panel C: Winner characteristics						
	Education	Criminal Record	Total Assets (Asinh)	General Caste	Female	Age
RD_Estimate	0.06	-0.00	-0.29**	0.01	-0.03**	-0.34
	(0.075)	(0.018)	(0.124)	(0.020)	(0.012)	(0.340)
	[0.397]	[0.851]	[0.021]	[0.744]	[0.019]	[0.311]
Dep var mean	2.516	0.133	12.849	0.280	0.430	41.111
Bandwidth	416	426	636	423	678	538
Effective Obs	14184	7244	10787	12611	19699	18880

This table reports regression discontinuity estimates from the main estimating equation of the effect of being above the treatment population threshold on electoral outcomes in that village's GP. This table restricts to the sample of elections that are not reserved for a particular caste category. The outcomes in Panel A are related to the overall political competition in the GP, panel B relates to the characteristics of candidates contesting for the position of pradhan, and panel C focused on the characteristics of the candidate elected as pradhan. The data is from the Uttar Pradesh State Election Commission for the 2015 and 2020 elections, including aggregate voting data and candidate affadavits. The outcomes in panel A are voter turnout, number of candidates per 1000 residents of the GP, effective candidates (receiving more than 5 percent of votes) per 1000 residents, the total number of candidates, a Herfindahl index of voting shares, and the winners' margin of victory. The outcomes in Panel B are the average education level, criminal record, age, and asset holdings (inverse hyperbolic sine) of candidates, as well as fraction of candidates who are general caste and female. The outcomes in Panel C are the education level, criminal record, age, asset holdings (inverse hyperbolic sine), caste identity and gender of the person elected pradhan. The running variable is the population of the village in the 2011 census. Each specification uses a linear polynomial, triangular kernel, and MSE-optimal bandwidth estimated following Calonico et al. (2017). Standard errors are clustered at the gram panchayat level and reported below the point estimates. p-values are reported within brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Heterogenous effects based on past GP composition

	Panel A: Largest village in previous GP							
	Index	Jan Dhan	BPL	Health insur.	Pensions	Saubhagya	LPG	PMMVY
RD_Estimate	0.06	5.72	3.08	-0.56	-1.96	2.61	12.09***	-0.15
	(0.055)	(3.743)	(4.450)	(1.790)	(1.837)	(2.932)	(4.650)	(0.234)
	[0.256]	[0.126]	[0.489]	[0.755]	[0.285]	[0.374]	[0.009]	[0.530]
Dep var mean	0.106	77.498	110.734	25.846	40.151	50.876	107.421	2.982
Bandwidth	246	218	325	373	384	299	263	372
Effective Obs	4639	4433	6413	7766	8076	5984	4572	7712
	Panel B: Not largest village							
	Index	Jan Dhan	BPL	Health insur.	Pensions	Saubhagya	LPG	PMMVY
RD_Estimate	0.10***	5.25**	8.40**	4.59***	-0.57	4.28*	3.59	0.35*
	(0.039)	(2.603)	(3.824)	(1.496)	(1.527)	(2.270)	(2.978)	(0.198)
	[0.008]	[0.044]	[0.028]	[0.002]	[0.709]	[0.059]	[0.227]	[0.075]
Dep var mean	0.105	76.727	110.189	25.876	40.134	50.844	106.044	2.964
Bandwidth	351	432	385	426	367	473	426	646
Effective Obs	9363	11921	11122	11024	10534	13020	12061	17155
	Panel C: Largest village in previous GP (NREGS)							
	Index	Work demand	Days worked	Labor exp.	Material exp.	Total projects		
RD_Estimate	0.01	0.00	0.02	2.46	-6.80	0.00		
	(0.048)	(0.003)	(0.106)	(19.450)	(9.830)	(0.003)		
	[0.785]	[0.268]	[0.830]	[0.899]	[0.489]	[0.737]		
Dep var mean	0.116	0.081	2.121	410.853	153.717	0.088		
Bandwidth	339	294	311	325	417	326		
Effective Obs	31035	27278	28945	29844	36916	29669		
	Panel D: Not largest village (NREGS)							
	Index	Work demand	Days worked	Labor exp.	Material exp.	Total projects		
RD_Estimate	0.28***	0.02***	0.55***	114.98***	48.33***	0.01***		
	(0.032)	(0.002)	(0.066)	(12.369)	(7.575)	(0.002)		
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]		
Dep var mean	0.110	0.080	2.103	407.160	153.541	0.088		
Bandwidth	510	572	549	567	449	440		
Effective Obs	64436	69614	68491	71018	55179	53912		

This table reports regression discontinuity estimates from the main estimating equation of the effect of being above the treatment population threshold on different welfare program and NREGS outcomes. The outcomes in Panels A and B are beneficiaries for welfare programs as measured in the Mission Antyodaya data in 2019, while Panels C and D investigate NREGS program outcomes. The panels split the sample based on whether the village was the largest village in its GP prior to 2015. The running variable is the population of the village in the 2011 census. Each specification uses a linear polynomial, triangular kernel, and MSE-optimal bandwidth estimated following Calonico et al. (2017). Standard errors are clustered at the gram panchayat level and reported below the point estimates. p-values are reported within brackets * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Effects on village amenities (bandwidth of 200)

Panel A: Educational outcomes

	2001		2011			2019
	Primary school	Middle school	Primary school	Middle school	Above Primary Ed	Education score
RD_Estimate	0.03**	0.02	0.02	0.06***	20.32*	2.28***
	(0.012)	(0.013)	(0.012)	(0.018)	(10.997)	(0.615)
	[0.019]	[0.136]	[0.125]	[0.001]	[0.065]	[0.000]
Dep var mean	0.864	0.162	0.864	0.493	693.339	33.448
Bandwidth	200	200	200	200	200	200
Effective Obs	15742	15742	15743	15743	15216	15612

Panel B: Village-level infrastructure

	2001		2011		
	Paved road	Electrified	Paved road	Electricity	FPS (=1)
RD_Estimate	0.03*	0.01	-0.01	0.91	0.03**
	(0.017)	(0.017)	(0.016)	(0.756)	(0.013)
	[0.062]	[0.692]	[0.520]	[0.231]	[0.025]
Dep var mean	0.612	0.300	0.718	21.659	0.810
Bandwidth	200	200	200	200	200
Effective Obs	15742	15742	15688	15741	15743

Panel C: Household-level infrastructure

	House			Sanitation		
	Brick	Organic	Piped water	Toilet	Open defec	Closed drains
RD_Estimate	1.46*	-1.21*	-0.02	3.28***	-3.41***	1.24**
	(0.774)	(0.730)	(1.002)	(0.681)	(0.706)	(0.540)
	[0.060]	[0.097]	[0.987]	[0.000]	[0.000]	[0.022]
Dep var mean	64.484	29.875	20.047	19.038	79.844	7.344
Bandwidth	200	200	200	200	200	200
Effective Obs	15741	15741	15741	15741	15741	15741

See the table note from Table 2 for further details. The running variable is the population of the village in the 1991 census. Each specification uses a linear polynomial, uniform kernel, and MSE-optimal bandwidth estimated following Calonico et al. (2017). Standard errors are clustered at the gram panchayat level and reported below the point estimates. p-values are reported within brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: Effect on delivery of services (bandwidth of 200)

	Programs Index	Jan Dhan	Below Poverty Line Card	Health insur-ance	Pension	Saubhagya electr.	LPG	Housing benefits	PMMVY
Panel A: Individual-level programs									
RD_Estimate	0.08***	5.26**	5.67*	2.45**	-1.31	3.82**	5.84**	5.60**	0.25
	(0.030)	(2.107)	(3.011)	(1.241)	(1.287)	(1.911)	(2.505)	(2.202)	(0.179)
	[0.007]	[0.012]	[0.060]	[0.048]	[0.308]	[0.045]	[0.020]	[0.011]	[0.169]
Dep var mean	0.107	76.887	110.675	25.856	40.219	50.980	107.011	45.816	3.003
Bandwidth	200	200	200	200	200	200	200	200	200
Effective Obs	16005	16005	16005	16005	16005	16005	16005	16005	16005
Panel B: Workfare program implementation									
	NREGS Index	Work demand	Days worked	Labor expend.	Material expend.	Total projects			
RD_Estimate	0.15***	0.01***	0.30***	60.34***	21.74***	0.01***			
	(0.031)	(0.002)	(0.065)	(12.339)	(6.854)	(0.002)			
	[0.000]	[0.000]	[0.000]	[0.000]	[0.002]	[0.004]			
Dep var mean	0.118	0.081	2.122	411.598	154.415	0.088			
Bandwidth	200	200	200	200	200	200			
Effective Obs	71033	71033	71033	70923	70923	70923			

This table reports regression discontinuity estimates from the main estimating equation of the effect of being above the treatment population threshold on different welfare program and NREGS outcomes. Panel A presents outcomes related to welfare programs as measured in the Mission Antyodaya data in 2019. See the table notes for 3 for details of the outcomes. The running variable is the population of the village in the 2011 census. Each specification uses a linear polynomial, triangular kernel, and fixed bandwidth of 200. Standard errors are clustered at the gram panchayat level and reported below the point estimates. p-values are reported within brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C Checking discontinuity usage by other programs

As with any regression discontinuity design, a major concern is that other variables may vary discontinuously across the cutoff, and so the estimates do not solely capture the effect of the treatment of interest. That concern is particularly salient here, where round population figures such as 1000 could plausibly be used by policymakers for other state or national-level programs. For example, Asher and Novosad (2020) study a national road-building program in which hamlets with populations of more than 500 or 1000 persons in the 2001 census were more likely to receive all-weather roads, Burlig and Preonas (2021) examine a national electrification program that targeted hamlets with populations greater than 300 people (2001 census), and Spears and Lamba (2016) evaluate a sanitation program with village-level incentives to build toilets based on the village population in the 2001 census.

Although we are unaware of any other programs that use population thresholds of 1000 persons in the 1991 and 2011 censuses, we implement five robustness checks and find this cannot explain the results. The first addresses the concern of all-India programs that uses the threshold. Note that the three papers listed in the previous paragraph studied national-level programs, so national programs are a particular concern. Panels (a) and (b) of Figure A12 plot the relationship between village population and GP population across all of the other states of India (similar to Figure 1). These figures demonstrate that there is not a discontinuity at a village population of 1000 persons in either the 1991 or 2011 censuses, confirming that other states did not use this cutoff for delimitation of GPs. We can thus check for the presence of national programs using this cutoff by checking to see if there is a discontinuity in the outcome variables in other states. If so, this would indicate that our results could be explained by a national program rather than GP delimitation.

Table A11 estimates the same regression discontinuity equation as in Table 2 in states other than Uttar Pradesh.[26] It finds no relationship with any of the study outcomes in those states, indicating that national level programs using a population discontinuity of 1000 people cannot explain our results. This is perhaps unsurprising given the range of different outcomes over which we observe effects: it seems unlikely that there would be a single national-level program that generates improvements across such a wide and differentiated set of public goods.

While this rules out national-level programs, Uttar Pradesh may have used this discontinuity for state-specific programs. Our second check uses a similar approach in checking whether there is a discontinuity in the outcomes in areas of Uttar Pradesh that did not follow the delimitation rules. In both 1995 and 2015, approximately 15% of Uttar Pradesh districts did not adhere to the rules. For these districts, there is no first relationship between village population and GP population in both years (panels (c) and (d) of Figure A12).[27] Table A12 shows that there is no discontinuity in

[26]We can only do this in the census data because we do not have the other outcome data for other states.

[27]We identify these districts by running the first-stage regression in each of them individually and selecting those with estimated t-statistics less than 1.5 or a first stage coefficient of less than 0.15

outcomes in those districts around the population threshold, inconsistent with other state programs using this eligibility cut-off. While it could be the case that these districts did not follow population-based rules for other programs, the non-compliant districts are mostly different in 1995 and 2015 – with a correlation of only 0.15 – suggesting that this non-compliance is idiosyncratic rather than reflecting fixed characteristics that make them less likely to follow rules for other programs.

The third check leverages the fact that prior to the 2015 delimitation, some villages were the lone village in their GP but still had populations around the cutoff. Since their GP could not be divided any further, the delimitation process did not affect them. However, villages in this group that were to the right of the cutoff would still have benefited from any state or national program using that cutoff for eligibility, so they form an ideal placebo sample. Figure A7 divides the sample by GP population prior to the 2015 delimitation (bins of 750-1500, 1500-2250, etc.), and re-runs the main specification within each bin. For all outcomes, there is no effect for villages who would have been the only village in their pre-2015 GP (bin of 750-1249).

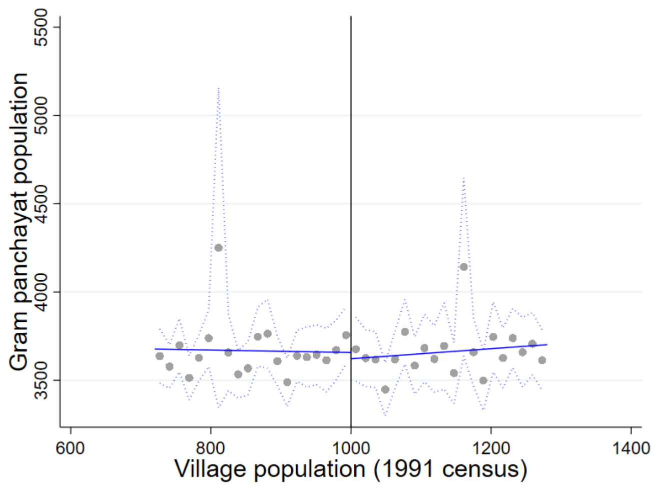
The fourth check takes advantage of how the delimitation of one village affects the other villages that were in the same panchayat prior to delimitation. Suppose that between 1995-2014, villages a and b were in a gram panchayat with village c, but village c's population is just above the population cutoff for the 2015 delimitation. Village c is more likely to be split into a GP of its own, meaning that villages a and b will be left in a GP with 1000 fewer people. However, villages a and b need not themselves have populations near the 1000-person threshold, so any effects for them cannot be the result of another state-specific program using that cut-off. To implement this design, for each GP that existed in 2011, we determine the village whose 2011 census population was closest to a population of 1000. We then drop those villages from the sample, but use their population as a running variable for the remaining villages in their GP. As seen in panel (a) of Figure A8, there is a strong discontinuity. Section 5.4 runs and discusses this analysis, showing that the estimated treatment effects are quite similar with this approach as in our main specification.[28]

As a fifth check, we examine the relationship between the outcomes and the village population in the 2001 census. This population was not used for delimitation of GPs (panels (e) and (f) of Figure A12), but might have been used for other programs if Uttar Pradesh has a propensity to use a population cutoff of 1000 for program eligibility. Table A13 does not detect any systematic pattern of discontinuities in 2011 related to a 2001 census population of 1000 people – one coefficient is statistically significant at the 10% level, as may be expected by chance. Putting these together, we find no evidence of alternative programs using the 1000 person discontinuity, and so conclude our estimates reflect polity size.

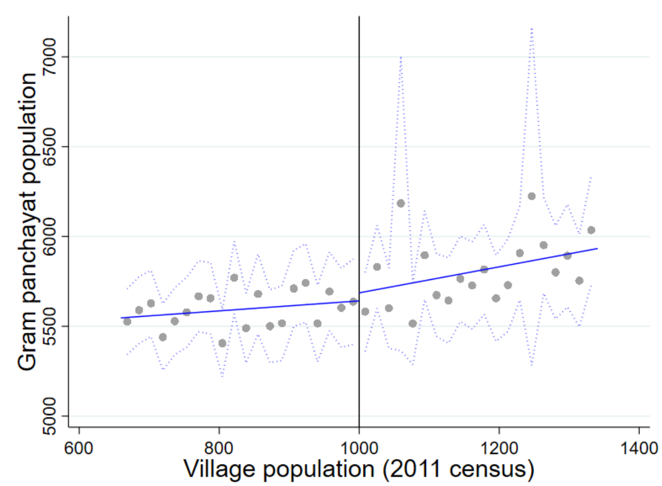
[28] This is also made apparent by the other analysis in section 5.4, which estimates heterogeneous treatment effects related to counterfactual GP size. We find the strongest effects when a village shifts from being in a relatively large GP to a solo-village GP. If this effect were simply about other programs that the village gets as a result of being above the discontinuity, we would not expect to see this pattern of heterogeneity.

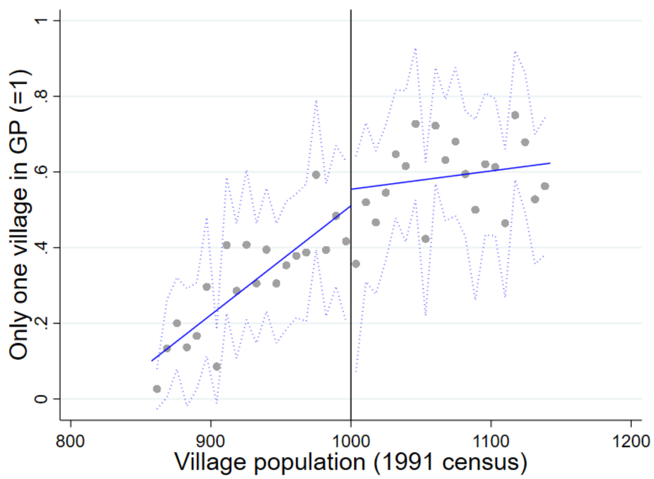
Figure A12: First stage of village population on gram panchayat population in placebo samples

(a) Other states (1991 census)

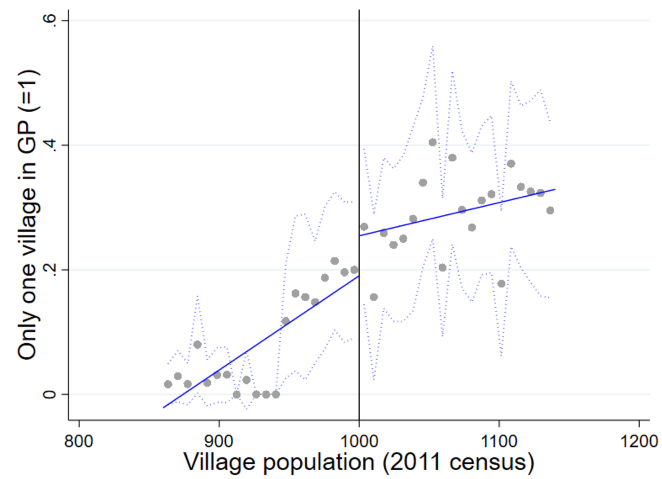


(b) Other states (2011 census)

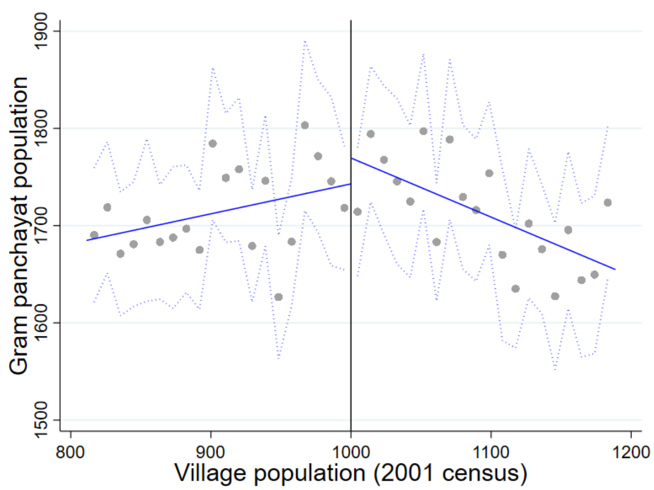




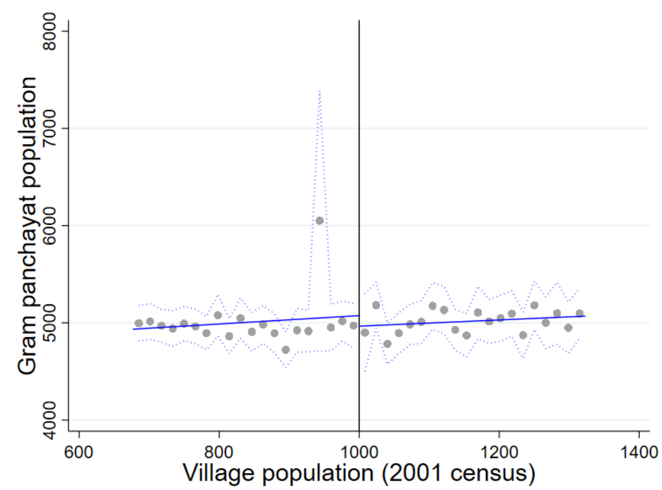
(c) Non-compliant districts (1991 census)



(d) Non-compliant districts (2011 census)



(e) Uttar Pradesh (2001 census population)



(f) Other state (2001 census population)

Notes: These figures plot the first stage relationship between a village's population in the relevant census and the population of the gram panchayat that a village is in. Sub-figures (a) and (b) are for states other than Uttar Pradesh. Sub-figures (c) and (d) are restricted to districts in Uttar Pradesh in which the delimitation was not implemented according to the official rules. Sub-figures (e) and (f) are using the 2001 village population in Uttar Pradesh and other states respectively as the running variables. They examine the relationship with the GP population immediately prior to the 2011 census census.

Table A11: Effects on amenities (other states)

Panel A: Educational outcomes						
	2001		2011			
	Primary school	Middle school	Primary school	Middle school		
RD_Estimate	0.00	-0.01	0.00	-0.01		
	(0.003)	(0.009)	(0.002)	(0.009)		
	[0.992]	[0.186]	[0.520]	[0.211]		
Dep var mean	0.958	0.393	0.985	0.655		
Bandwidth	327	333	228	256		
Effective Obs	65776	64111	44367	52457		
Panel B: Village-level infrastructure						
	2001		2011			
	Paved road	Electrified	Paved road	Electricity	FPS (=1)	
RD_Estimate	0.01*	0.01	0.01	0.51	0.01	
	(0.008)	(0.009)	(0.008)	(0.567)	(0.011)	
	[0.083]	[0.245]	[0.118]	[0.372]	[0.266]	
Dep var mean	0.611	0.727	0.724	56.987	0.713	
Bandwidth	385	314	327	347	287	
Effective Obs	71448	43840	60738	72140	57446	
Panel C: Household-level infrastructure						
	House		Sanitation			
	Brick	Organic	Toilet	Open defec	Closed drains	In house water
RD_Estimate	-0.00	0.23	0.17	-0.12	0.19	0.37
	(0.471)	(0.439)	(0.466)	(0.492)	(0.139)	(0.418)
	[0.995]	[0.604]	[0.717]	[0.812]	[0.169]	[0.376]
Dep var mean	32.632	50.069	26.273	71.976	3.725	28.476
Bandwidth	346	453	336	334	428	453
Effective Obs	71470	96092	69818	68817	89557	95732

This table reports regression discontinuity estimates from the main estimating equation of the effect of being above the treatment population threshold on different public-service related outcomes for states other than Uttar Pradesh. Panel A presents outcomes related to education: whether the village had a primary or middle school present in 2001 or 2011 (as measured in the census), the number of residents with a primary education or above in 2012 (Socio-economic and Caste Census), and the education score of the village in the 2019 Mission Antyodaya survey. The outcomes in panel B are whether the village has an all-weather road in 2001, whether the village has electricity in 2001, whether the village has an all-weather road in 2011, the proportion of households within the village who had electricity in their home in 2011, and whether the village has a Fair Price Shop within the village in 2011 (all measured in the respective Census round). The outcomes in panel C are the fraction of village residents living in houses made of brick or houses made of organic materials (mud, etc.) in 2011, having in-house piped water in 2011, having a toilet in 2011, primarily defecating in the open in 2011, and with closed drain sanitation systems. See the online Appendix for details of each data set. The running variable is the population of the village in the 1991 census. Each specification uses a linear polynomial, triangular kernel, and MSE-optimal bandwidth estimated following Calonico et al. (2017). Standard errors are clustered at the

Table A12: Effects on village amenities (placebo districts)

	Panel A: Educational outcomes					
	2001		2011		2019	
	Primary school	Middle school	Primary school	Middle school	Education score	
RD_Estimate	-0.00	0.04	0.01	0.01	0.12	
	(0.031)	(0.032)	(0.028)	(0.049)	(1.707)	
	[0.958]	[0.257]	[0.808]	[0.764]	[0.944]	
Dep var mean	0.865	0.155	0.856	0.488	33.262	
Bandwidth	309	521	535	396	394	
Effective Obs	1570	2999	3212	2131	2130	
	Panel B: Village-level infrastructure					
	2001		2011			
	Paved road	Electrified	Paved road	Electricity	FPS (=1)	
RD_Estimate	0.02	0.06	0.03	0.62	0.04	
	(0.035)	(0.041)	(0.040)	(1.812)	(0.031)	
	[0.548]	[0.146]	[0.423]	[0.730]	[0.252]	
Dep var mean	0.600	0.296	0.709	21.794	0.797	
Bandwidth	680	479	529	531	466	
Effective Obs	4105	2829	3124	3150	2764	
	Panel C: Household-level infrastructure					
	House		Sanitation			
	Brick	Organic	Toilet	Open defec	Closed drains	In house water
RD_Estimate	-2.42	2.49	2.10	-2.03	0.24	-1.06
	(2.805)	(2.867)	(1.492)	(1.420)	(1.382)	(2.353)
	[0.388]	[0.385]	[0.159]	[0.152]	[0.860]	[0.651]
Dep var mean	64.356	29.890	18.521	80.594	7.134	42.768
Bandwidth	409	381	515	651	453	464
Effective Obs	2328	2116	3092	3769	2763	2669

This table reports regression discontinuity estimates for districts in Uttar Pradesh that do not follow the delimitation rules from the main estimating equation of the effect of being above the treatment population threshold on different public-service related outcomes. Panel A presents outcomes related to education. The outcomes in panel B are whether the village has an all-weather road in 2001, whether the village has electricity in 2001, whether the village has an all-weather road in 2011, the proportion of households within the village who had electricity in their home in 2011, and whether the village has a Fair Price Shop within the village in 2011 (all measured in the respective Census round). The outcomes in panel C are the fraction of village residents living in houses made of brick or houses made of organic materials (mud, etc.) in 2011, having in-house piped water in 2011, having a toilet in 2011, primarily defecating in the open in 2011, and with closed drain sanitation systems. See the online Appendix for details of each data set. The running variable is the population of the village in the 1991 census. Each specification uses a linear polynomial, triangular kernel, and MSE-optimal bandwidth estimated following Calonico et al. (2017). Standard errors are clustered at the gram panchayat level and reported below the point estimates. p-values are reported within brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: Effects on amenities (2001 population discontinuity)

Panel A: Educational outcomes

	2001		2011		2019
	Primary school	Middle school	Primary school	Middle school	Education score
RD_Estimate	-0.00	0.01	0.01	0.01	-0.07
	(0.012)	(0.009)	(0.012)	(0.012)	(0.367)
	[0.920]	[0.403]	[0.383]	[0.573]	[0.850]
Dep var mean	0.801	0.123	0.826	0.391	31.179
Bandwidth	356	477	374	650	658
Effective Obs	19089	26251	18683	33792	36100

Panel B: Village-level infrastructure

	2001		2011		
	Paved road	Electrified	Paved road	Electricity	FPS (=1)
RD_Estimate	-0.01	0.01	0.01	0.55	-0.02*
	(0.013)	(0.010)	(0.012)	(0.502)	(0.012)
	[0.594]	[0.521]	[0.500]	[0.270]	[0.097]
Dep var mean	0.569	0.277	0.683	21.668	0.727
Bandwidth	472	654	499	677	446
Effective Obs	26251	36298	27039	35892	23606

Panel C: Household-level infrastructure

	House		Sanitation			
	Brick	Organic	Toilet	Open defec	Closed drains	In house water
RD_Estimate	0.30	-0.61	0.10	0.02	-0.04	-0.18
	(0.591)	(0.489)	(0.565)	(0.598)	(0.408)	(0.810)
	[0.615]	[0.212]	[0.856]	[0.975]	[0.913]	[0.828]
Dep var mean	63.536	30.854	17.281	81.560	7.116	41.810
Bandwidth	525	688	354	402	491	389
Effective Obs	29071	38368	19253	19338	20731	21721

This table reports regression discontinuity estimates for the main estimating equation of the effect of being above the treatment population threshold on different public-service related outcomes. Panel A presents outcomes related to education. The outcomes in panel B are whether the village has an all-weather road in 2001, whether the village has electricity in 2001, whether the village has an all-weather road in 2011, the proportion of households within the village who had electricity in their home in 2011, and whether the village has a Fair Price Shop within the village in 2011 (all measured in the respective Census round). The outcomes in panel C are the fraction of village residents living in houses made of brick or houses made of organic materials (mud, etc.) in 2011, having in-house piped water in 2011, having a toilet in 2011, primarily defecating in the open in 2011, and with closed drain sanitation systems. The running variable is the population of the village in the 2001 census. Each specification uses a linear polynomial, triangular kernel, and MSE-optimal bandwidth estimated following Calonico et al. (2017). Standard errors are clustered at the gram panchayat level and reported below the point estimates. p-values are reported within brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D Spillovers

A concern with the interpretation of the results is the possibility of financial spillovers between communities, and in particular, whether the increase in benefits in villages allocated to a smaller GP come at the expense of less spending in untreated villages. In such a zero-sum world, our findings would be just as informative about the mechanisms underlying the effects of polity size, showing that leaders in smaller polities exert greater effort in service delivery. However, the policy implications would be more ambiguous and require consideration of distributional effects. A similar concern is present with any policy executed at a policy-relevant scale, such as business training/ loans/ individual job training that may take business from untreated businesses/job seekers rather than expand overall output (Crépon et al., 2013) or other papers measuring performance through public goods access (e.g. Martinez-Bravo, 2017). A nice of feature of our context is that it is possible to test for since the boundaries of spillovers are well-defined.

The key to whether or not negative spillovers occur is whether there are hard budget caps on funding for the relevant item *and* whether those caps bind in practice. If either condition does not hold – budget constraints are soft or additional spending does not push spending into the cap (and so does not require a loss for others) – then budgetary gains for one GP do not imply losses for another. This may be plausible in this environment, especially given that even an upper bound on the increased costs from the delimitation policy are small in the broader context – for example, an increase in the number of GPs by 10% due to the rule multiplied by the largest treatment effect we estimate would only induce an increase in spending in that category of around 1%. We take two broad approaches to understanding spillovers: first, examining budget allocations to assess whether conditions exist for spillovers; and second, testing empirically for spillovers.

We first examine whether budget caps bind by examining whether annual budgets are typically fully spent. Between 2004 and 2009, the Comptroller Audit General of India (CAG) conducted audits on a sub-sample of district, block and village panchayats to determine how effectively the funds were being utilized. This survey was intended to be representative of the state, and although the exact numbers varied year to year, it covered an average of 20 district panchayats, 60 blocks, and 1000-2000 GPs. District panchayats receive the bulk of the financing that passes to rural areas from the Ministry of Rural Development. They are the main source of determination for fund allocation to the lower tiers, and help implement and drive many of the anti-poverty programs set up by the Ministry of Rural Development. At the district (*zila*) level, Figure A13 finds a notable level of under-utilization of funding in each year. Significant portions of the budgets between 2006-2009 also remain under-utilized at the block (*kshetra*) and GP levels. If budgets are not being fully spent down each year, then additional funding needs of the magnitude induced by the treatment we study appear feasible to absorb within the slack capacity. However, this might not be true for a substantively larger decentralization push, a point that we will return to below.

Figure A13: Utilization of funds across the three tiers

(₹ in crore)

Year	Number of PRIs checked	Opening balances	Funds received	Total funds received	Expenditure (per cent in bracket)	Closing balances
Zila Panchayats						
2006-07	52	338.56	476.91	815.47	497.80(61)	317.67
2007-08	52	319.41	589.80	909.21	484.00(53)	425.21
2008-09	55	439.04	993.15	1432.19	1022.87(71)	409.32
Kshetra Panchayats						
2006-07	139	51.19	160.57	211.76	151.53(72)	60.23
2007-08	130	53.33	282.39	335.72	274.59(82)	61.13
2008-09	300[3]	156.36	532.09	688.45	503.09(73)	185.36
Gram Panchayats						
2006-07	2430	39.18	135.36	174.54	132.32(76)	42.22
2007-08	4525	87.28	376.92	464.20	346.73(75)	117.47
2008-09	3003[4]	71.85	363.89	435.74	307.84(71)	127.90

Notes: This table is taken from the Comptroller and Auditor General of India's audit of Gram Panchayats in Uttar Pradesh in 2009. The table is in crores (a unit measuring 10 million) and in rupees. The data comes from the audit of a representative sample of GPs in the state of Uttar Pradesh.

D.1 Measuring spillovers in practice

The previous section found that the conditions may be present for a lack of negative spillovers. This section follows up on this by directly testing for negative spillovers in the outcomes we study. The concern that we wish to test for is that if one local government has a larger number of other local governments with which it is potentially competing for resources, does this worsen access to services for its populations? We assess this at the block level, as many development related programs are administered at this level and so this is a likely place at which negative spillovers could emerge.

The ideal experiment would be one in which the average GP size is randomized at the block level, producing cross-sectional variation in the density of political units across blocks and accounting for spillovers or other general equilibrium effects at the block-level. We approximate this experiment in our data by again leveraging the discontinuity in GP size. We use variation coming from whether the block has a larger fraction of villages just above the threshold as opposed to below: blocks with a higher proportion of villages above the cutoff will have a larger number of GPs, but are plausibly otherwise similar to those with a higher proportion just below. More precisely, for each village, we examine the villages in their block with a population of between 900 and 1100 people. Leaving out the village itself, we calculate the fraction within the same block and this bandwidth that are above the population cutoff of 1000 ($frac/_{above}$) and run the following regression:[29]

[29] Results are similar with bandwidths of 50 and 200 people around the cut-off.

$$Y_v = \beta_0 + \beta_1 frac\_above_v + \gamma X_v + \epsilon_v \tag{2}$$

where Y_v is the outcome of interest, X_v is a set of village-level controls, and standard errors are clustered at the block level. The first column of panel A tests whether having a higher fraction of villages above the cut-off is indeed related to the total number of GPs in the block. We find that it is ($p < .001$): on average, there are 7.4 more GPs among the blocks where all the villages were above the threshold as compared to those with $frac/_{above}$ equal to zero. Since the average block has a total of 69.2 GPs, this shock could have a moderate impact on intra-GP competition for resources.

One concern is that the fraction of villages above the cutoff could plausibly be related to other characteristics of the block that determine public goods delivery and so the estimates in the above equation could be biased. The remaining columns of panel A conduct placebo checks for whether the fraction of villages above the cutoff is related to village characteristics in the 1991 census. We do not find any evidence of this, suggesting that this approach can be used to test for spillovers. This is particularly encouraging given that our primary outcomes are exactly these variables, but measured in later census round, so there should not be pre-existing imbalances on those variables.

In panels B and C, we do not observe statistically significant negative spillovers for any of the outcomes. We focus on outcomes for which there were statistically significant results in the main analysis, as those are the cases in which negative spillovers could occur. In two cases, the coefficients are even statistically significant, but *positive* (middle schools and toilet construction) rather than negative. This surprising result could be due to chance or potentially come from yardstick competition, where villages see elevated public goods delivery in their neighboring villages and so demand greater services; we test for this in another project using a different empirical strategy. But more broadly, these estimates point against large negative spillovers.

The lack of large negative spillovers is consistent with the moderate-sized point estimates that we observe for different public goods and modest variation in the number of GPs induced by this instrument. For example, in most cases, the instrument is related to a difference of only 2-4 GPs at the block level, and multiplied by our estimated treatment effects, this would only have a modest effect on block-level expenditures. As a result, if there is some give in the budget constraint, we would not expect large negative spillovers. However, if the government were to undertake a delimitation exercise that produced significantly more local government units such as doubling the total number of GPs, then negative spillovers will be more likely to occur due to reaching the budget constraint. However, for more limited exercises, like the episode we study, the overall effect on public goods delivery is positive.

Table A14: Block-level spillovers

Panel A: Placebo checks (values in 1991 census)							
	First stage	Placebo checks (1991)					
	Total GPs	Primary school	Middle school	Paved road	Electrified	Literate pop	SC pop
Frac above cutoff	8.10***	0.02	0.01	0.04	0.01	-18.54	-0.02
	(1.767)	(0.020)	(0.012)	(0.029)	(0.041)	(15.201)	(0.014)
Dep var mean	69.232	0.588	0.118	0.421	0.191	307.216	0.244
Observations	94058	94927	94927	94927	94470	94927	94927

Panel B: Educational outcomes						
	2001		2011			2019
	Primary school	Middle school	Primary school	Middle school	Above Primary Ed	Education score
Frac above cutoff	0.01	-0.00	0.03	0.04*	-13.72	0.16
	(0.015)	(0.013)	(0.021)	(0.021)	(36.393)	(0.803)
Dep var mean	0.689	0.182	0.718	0.409	735.814	32.300
Observations	94737	94737	94926	94926	90041	93822

Panel C: Other outcomes							
	Road		Electricity		Household (2011)		
	2001	2011	2001	2011	Brick wall	Toilet	Closed drains
Frac above cutoff	0.02	0.03	-0.03	-3.36	2.37	4.92**	-0.26
	(0.026)	(0.042)	(0.042)	(2.610)	(3.297)	(2.005)	(0.793)
Dep var mean	0.596	0.660	0.305	23.397	63.066	18.328	6.801
Observations	94737	94583	94737	94867	94867	94867	94867

Panel D: Post-2015 outcomes			
	First stage	Outcomes	
	Total GPs	Programs Index	NREGS Index
Frac above cutoff	4.82**	-0.07	-0.02
	(2.037)	(0.062)	(0.097)
Dep var mean	78.816	0.000	0.000
Observations	104619	101554	457920

This table reports reduced form estimates of the spillovers of delimitation on nearby villages. We include controls for the total number of villages in the block and the village's own population. Standard errors are clustered at the gram panchayat level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.