Appendix for *Citizenship, Inequality, and Urban Governance*: Findings from Ahmedabad, Bhavnagar, Kochi, Mumbai, and Vadodara

Appendix 1: Note on Weighting

Our survey collected a range of demographic information from respondents, including gender, religion, and caste group. In addition, the survey enumerators were tasked with identifying the housing type of each respondent's dwelling. Dwellings were categorised as one of five types: informal shack settlements, informal slum settlements, lower middle class, upper middle class, or upper class. In the sample, those belonging to the first level of housing type (informal shacks) were deliberately oversampled. This included a "booster" sample which was in addition to the original, randomised sample.

We have chosen to reweight the sample data according to respondents' housing type. The type of dwelling is here conceptualised as a proxy for socioeconomic status. It is also expected to correlate with caste status. For housing type, we use novel benchmark data collected as part of the project. This household listing data, collected from March 2019 to June 2019 in 7 project cities, provides a comprehensive, census-like account of the distribution of dwelling types in each city.

On the other hand, we have chosen not to weight based on caste and religion since there are no census numbers for caste (or religion) categories other than SC/ST, whereas our survey gave respondents the option to identify as Forward/General castes, OBCs, or as no caste at all (in addition to SC or ST) (Census 2011). In general, weighting necessitates that there are reliable population margins for all categories of a variable upon which one seeks to adjust one's data (Solon et. al. 2015). Since we lack reliable population margins for OBCs or General/Forward castes, we cannot adjust our data to correct for over or undersampling. Accordingly, reweighting along caste would necessitate that we collapse our caste categories into SC, ST, and a very large "Other" umbrella. Doing so would likely bias our reported results, particularly for the portion of respondents not belonging to SC/ST groups. We did not weight on gender because the focus of the survey is on the household, rather than individual level, and so should not greatly affect results. Empirically, religion and gender weights were shown not to significantly affect the reported results for a representative subset of the questionnaire.

We have reason to believe that reweighting along housing type mitigates the effect of higher proportions of Dalits and Adivasis. From a theoretical perspective, housing type serves as a reasonable proxy for socioeconomic status. Given the relatively poor economic condition of many individuals belonging to SC/ST groups, we expect that weighting along housing type will reduce bias introduced through larger proportions of this subpopulation (Bharathi et. al. 2019; Thorat et. al. 2015; and Vithayathil and Singh 2012).

We used the method of rake weighting (or iterative proportional fitting) to create weights that are unique to each city. For the seven cities (except Mumbai), each response is assigned a weight according to housing type, which are unique for each city according to the difference between the sample margins and the population distributions of the five housing types in the city (as determined by the listing data) (Kolenikov 2014). The frequencies and indices for Mumbai

presented here are unweighted. This is because we did not undertake a booster sample of HT1s in Mumbai, since the listing of household types yielded sufficient numbers of informal housing types. These weights are summarised in the table below:

	Weight Factor by Housing type				
City	**HT1 – Informal shack**	**HT2 –Informal slum**	**HT3 - Lower Middle Class**	**HT4 - Upper Middle Class**	**HT5 - Upper Class**
Bhavnagar	0.0052	0.3280	0.3784	2.8887	1.5862
Vadodara	0.0363	0.4267	0.1801	2.7497	1.4822
Ahmedabad	0.1029	0.7273	0.4731	2.1092	1.2591
Chennai	0.2427	0.6146	1.4697	1.0094	0.2294
Hyderabad	0.1608	0.9919	1.3834	0.6688	0.7045
Kochi	0.0025[1]	0.2189	.2918	2.4469	2.5371

[1] The listing data for Kochi returned zero HT1s. In order to be able to calculated raked weights, HT1s in Kochi were assigned an arbitrarily low population proportion of 0.0001, whereas the sample proportion of HT1s in Kochi was 0.0395

Appendix 2: Citizen Participation Index

The citizen participation index has three components: (1) electoral participation, (2) non-electoral political participation, and (3) civic participation. Electoral participation - voting in national, state, and municipal elections is coded 1 if a respondent voted in an election and 0 otherwise. This component averages voting across the three levels and is rescaled from 0 to 1; (b) non-voting participation includes whether a respondent is a party member, contributes time during election campaigns, attends political rallies and meetings between elections, and discusses specific candidates among family, friends and others within the community). Each of these elements takes the form of a dummy variable and is coded 1 for "yes" and 0 if "no". This component averages the above forms of non-electoral participation and ranges from 0 to 1; and (c) civic participation that includes whether a respondent attends ward committee meetings, holds membership in non-political, non-government organisations and associations, and perceptions of community participation in preventing harassment of women in the neighbourhood. Each of these is coded 1 if "yes" and 0 if "no". As with the above this component is also averaged and ranges from 0 to 1.

The citizen participation index is an additive index that weights all components of participation equally. The questions used in the index are as follows:

Sub-index type	Q No.	Question
Electoral Participation	A13	Have you voted in the recent Municipal Corporation elections?
	A14	Have you voted in the State Assembly Elections?
	A15	Have you voted in the Lok-Sabha Elections, (MP) 2014?
Non-Voting Political Participation	A20	Have you or has anyone in your household contributed time to election campaigns during any elections?
	A21	Have you or has anyone in your household participated in meetings/ rallies organised by political parties/ officials between elections?
	A22	Have you or has anyone in your household discussed supporting a candidate with friends, neighbours or other people in the community during elections?
	A23	Are you a member of a political party?
Civic Participation	A25a	Are you aware of a ward where local issues are discussed? If yes, in the last year, did you, or someone in your household, attend ward meetings?
	A26	Do you or a member of your household participate in any of the following organisations?
	A28	Do you think your neighbours will help (either in person or by making a phone call to the police or other neighbours) if they see a woman being harassed in the neighbourhood?

Higher values on the index represent higher levels of participation across the range of activities defined as participation. The value zero (0) represents the case where a citizen does not

participate in any activity - does not vote in any election at any level, does not participate in campaigning, either formally or informally, or attending rallies and so on, and does not engage in civic participation.

Conversely, the value one (1) represents the case where a citizen participates in all activities - votes in elections to all levels of government, participates in election campaigns and rallies, and in contributing resources, and attends ward committee meetings and engages in all forms of civic participation. While the end-points of the index mark the two extremes of citizen participation - no participation to full participation, and are clear to understand, the values in-between represent different combinations of the three components of participation and require some elaboration. More specifically, all other values been 0 and 1 represent either a balanced but incomplete level of participation (i.e. respondent participates in at least 1 activity in all three types of participation activities) or an unbalanced but complete level (i.e., a respondent participates fully in one set of participation activities but not in of the other forms of participation). For instance, a score of 0.33 on the index - considered a low level of participation - implies that a respondent either (a) only votes (in all elections at all three levels) *or* (b) participates in only non-voting forms *or* (c) participates only in civic activities. This score can also represent a combination where a respondent participates in fewer activities across the range of activities, that is, say votes in the Lok Sabha election, *and* attends only political rallies among the set of non-voting political activities, *and* only participates in a non-government organisation among the set of civic participation activities. The same logic can be extended while interpreting higher values of the participation index.

A score of, say, 0.5 suggests a higher level of participation - a respondent now participates in a greater number of activities across multiple forms of participation. That is, a respondent is likely to vote in all three elections *and* participate in one or more forms of other types of participation (say, (a) only participate in rallies and contribute resources between elections *or* (b) only participate in civic activities such as attending a ward committee meeting or membership in a caste-religious non-political organisation *or* (c) contributes resources (non-voting participation) *and* is a member of a civic organisation (civic participation). Alternatively, a respondent is likely to vote more (say in two elections) *and* engage in multiple civic activities (attend ward meetings and be a member of a regional society) but participate in fewer non-voting political activities. Or vote less but participate more in non-voting and civic participation.

Once again, note that as the score increases the possible combinations of the indicators of participation across all three types also increase. Consider a score of 0.66. Here, a respondent fully engages in only two types of participation *or* in at least two participation activities across all three forms. Values above 0.67 indicate that a respondent now fully engages in any two forms of participation and participates in an increasing number of activities in the third form of participation. That is, the number of activities in the third form of participation increases as the participation score goes beyond 0.67. Finally, the highest score of 1 indicates balanced and complete participation across all three components of participation.

Appendix 3: Basic Service Delivery and Infrastructure Index

The Basic Service Delivery and Infrastructure Index (BSDII) is based on 5 dimensions of household infrastructure: (i) water, (ii) sanitation, (iii) electricity, (iv) flooding/water-logging, and (v) type of road outside the house. The following sections detail the construction of each dimension of household infrastructure and the BSDII as well as the questions and the coding rules used in this process.

1. Water

The water infrastructure component of BSDII comprises of: (1) access to water, (2) convenience in accessing water, (3) the usability of the water a household receives, (4) the ability (and methods) of households to store water, and (5) the frequency and duration of water supply for a household. In order to measure these five components of household water infrastructure, we the following questions:

i. Access: What is the main source of water for members of your household? The response categories are: (1) Tap (piped water), (2) Well (covered or uncovered), (3) Hand pump, (4) Borewell, (5) other source. The variable is coded 1 if the water source is piped and 0 for all other sources.
ii. Convenience: Is this source of water inside or outside your home premises? The response categories are: (1) Inside and (2) Outside. This variable is coded 1 if the source of water is inside and 0 if the source is located outside the household.
iii. Usability: Is this (primary) source of water for general use or for drinking? The response options are: (1) General use (2) Drinking, (3) Both. This variable is coded 1 if the responses are either drinking or both and 0 if the response is general use only.
iv. Water Supply and Storage: We combine separate questions on the frequency and duration of water supply to households and the ability (and means) of households to store water over a period of time to generate a composite water supply and storage component.

We use two questions to measure the frequency and duration of water supply in household:

i. How many days a week does your household receive water supply? The response options range from 1 to 7.
ii. For how many hours in a day does your household receive water supply? The response options range from 0 to 24 hours.

We code the variable as shown:

	1-2 Hours	3-4 Hours	5-8 Hours	9-12 Hours	13-24 Hours
0 Days = 0					
1-3 Days	1	2	3	3	3
4-6 Days	3	4	5	5	5
7 Days	6	7	8	8	10

In order to measure the ability of a household to store over a period of time, we use the following two questions:

i. Do you have a water storage system for this source? The response options are (1) No, (2) Yes.
ii. What is the type of water storage system? The response categories are: (1) Movable containers (small sized), (2) Drum (medium sized), (3) Large Tank/Drum without motorised pump, (4) Large Tank/Drum with motorised pump, and (5) other.

Water storage is coded 0 if the household has no water storage system, 1 if a household has small movable containers, 2 if the household has medium sized containers, 3 in case of large sized storage drums, and 4 if the household has a large sized motorised storage system such as an overhead tank.

We follow two steps to create the water supply and storage component of water infrastructure. All households that neither receive water supply nor have storage are coded as 0 (that is, receive water supply on 0 days and do not have a storage system for water). All households that receive uninterrupted water supply (between 13 to 24 hours of water supply every day of the week) are coded 10. We then multiply the water supply variable with the water storage variable to arrive at a measure that combines the frequency and duration of water supply with the ability of a household to store water. For instance, a household that receives 1-2 hours of water 1 to 3 days a week and has small movable containers for water storage is coded 1. A household that receives 1-2 hours of water 1 to 3 days a week and has an overhead water tank with a motorised water pump for water storage is coded 4. A household that receives 3-4 hours of water 4 to 6 days a week and has an overhead water tank with a motorised water pump for water storage is coded 12.

The water infrastructure variable is an additive combination of the above five components. We rescale the variables so that the final water infrastructure variable is bound between 0 and 1. A value of 0 suggests that the household has a non-piped water source outside the household, with no water storage and no regular water supply. A value of 1 on the other hand indicates that a household has a piped water source located inside the household with a motorised overhead water storage system and an uninterrupted supply of water. Higher values on the water infrastructure variable indicate that a household has a relatively higher level of water infrastructure.

2. Power (Electricity):

We base the electricity infrastructure dimension of BSDII on three elements of electricity supply. First, does a household have power? Second, how often are there power outages in a week, and third, how many hours does a household go without power during such outages. The questions are as follows:

i. Does the house have electricity? The response options are: (1) No and (2) Yes
ii. How many times a week are there power cuts? The response options are: (1) Never, (2) Once a week, (3) Twice a week, and (4) More than twice a week.
iii. During power cuts, how many hours are you without electricity? The response categories are: (1) less than 1 hour, (2) 1-2 hours, (3) 2-6 hours, (4) 6-12 hours, and (5) More than 12 hours.

In the final power infrastructure variable constructed from the above three questions, a household that does not have power supply is coded 0. A household that has power supply and never faces power outages is coded 7. Rest of the households are coded based on the combination of frequency and duration of power outages as illustrated in the following table:

	Less than 1 hour	**1-2 Hours**	**2-6 Hours**	**6-12 Hours**	**More than 12 Hours**
No power cuts = 7					
Once a Week	6	6	5	4	3
Twice a Week	6	5	4	3	2
More than twice a Week	5	4	3	2	1
No Power = 0					

We rescale power infrastructure so that it ranges between 0 and 1. A value of 0 indicates that the household receives no power, and a value of 1 indicates a household with uninterrupted power supply.

3. Sanitation (Toilets and Sewerage):

In order to measure sanitation infrastructure in a household we consider the following elements: the type of toilet a household has and (among those with a modern sanitation system connected to a main sewer line) whether, and how often, the line connecting the household to the main sewer gets blocked. We construct this variable in two steps. In Step 1, we use the following question:

i. What kind of toilet facility does the household have? The categories are: (1) Flush/Pour Flush Latrine within premises connected to a Piped Sewer System, (2) Flush/Pour Flush Latrine within premises connected to Septic Tank, (3) Flush/Pour Flush Latrine within premises connected to any Other System (not connected to a sewer line) such as open drainage into ground or into water body through covered drain or uncovered drain, (4) Pit Latrine: Ventilated or Covered Pit, (5) Pit Latrine: Open Pit, (6) No Latrine within Premises: Public Latrine, and (7) No Latrine within Premises: Open Defecation.

This variable is coded 0 if the household has no toilet within premises and household members defecate in the open, 1 if the household has no toilet within premises but uses a public toilet, 2 if the toilet type is a pit (covered or uncovered), 3 if the household has a flush or pour toilet within premises but without a piped sewer connection, and 4 if the household has a flush or pour toilet connected to a septic tank or a piped sewer system.

We do this for all households in the sample. In order to further differentiate infrastructure in those households that have a flush/pour toilet connected to a septic tank or a piped sewer system (households with a toilet type 4 from the above coding rule) from other households (with toilet types 0 through 3) we use two additional questions as follows:

(ii) Does the sewer line near or in your house get blocked? The response options are (1) No, and (2) Yes

(iii) How often does this happen? The response categories are: (1) Very rare, (2) Once a year, (3) 2-3 times a year, (4) Once a month, and (5) More than once a month. In Step 2, we recode these questions as 0 for a household that experiences blockages once or more in a month, 1 if the sewer gets blocked 2- 3 times a year, 2 if this happens once a year or very rarely, and 3 if a household never experiences blockages in the sewer line. Note that the above two questions are used only for the households that have a toilet type- 4.

The two variables from steps 1 and 2 are combined to create the final sanitation infrastructure variable. The sanitation infrastructure variable ranges from 0 (a household with no toilet, resorting to open defecation) to 5 (household with a flush/pour toilet connected to a piped sewer system or septic tank that experiences no blocks in the sewer line). Also note that a household that receives a value 4 in the sanitation infrastructure variable is one that is characterised by a flush/pour toilet connected to a septic tank or a piped sewer system but experiences blockages once or more in a month. As with the above, sanitation infrastructure is also rescaled to values between 0 and 1.

4. Water Logging/Flooding:

In order to measure the vulnerability of households to flooding and water logging during monsoon we ask the following question:

i. Does the ground floor of the building you live in get flooded during monsoon? The responses options are (1) Never, (2) Sometimes, and (3) Always. This variable is coded 0 for households that always face flooding during monsoon, 1 for households that sometimes experience flooding, and 2 for households that never get flooded. We then rescale it so it takes values between 0 and 1.

5. Roads:

We measure the type of road - whether paved (*pucca*) or unpaved (*kuccha*) using the following question:

(i) What is the type of road in front of your house? The response options are: (1) *Pucca*, and (2) *Kucchha*. We code this *kuccha,* and 1 if the road type in front of the household is *pucca.*

The BSDII is an additive index that includes the five components outlined above: water, power, sanitation, flooding, and road infrastructure. We construct BSDII by assigning equal weights to the three components that are directly connected to household infrastructure - water, power, and sanitation, and half-weights to flooding and roads. Implicit is the notion that the first three components "count" more for a household than the latter two. While we do not deny the importance of household flooding during monsoon and the condition of the roads in overall infrastructure, we argue that they are less salient to a household than are water, power, and sanitation.

The formula for our services index is given by:

BSDII = [(Water) + (Power) + (Sanitation) + 0.5*(Flooding) + 0.5*(Road)]/4

Appendix 4: Housing Type details

Category 5 Dwellings (HT-5): Upper-Class Housing

Picture-1: HT5

Independent house or apartment building

- Often constructed using materials in addition to concrete such as: glass, wood.
- If a house, multiple rooms, one family or joint family lives there. Generally, not multiple. independent units of unrelated families within one house. Can assess this by a single mailbox on the outside, single address marked doorway entrance.
- Usually has surrounding wall with gate in front of house.
- If apartment building will also have wall and gate with security guarding entrance.
- Often apartment complexes/gated communities. Amenities such as a swimming pool, shopping mall, gym, will be inside of complex.
- Size of individual apartments will be large.
- Multiple balconies for one apartment.
- Large windows.

Category 4 Dwellings (HT-4): Middle-Class Housing

- Independent house or apartment building.
- If independent house and large (more than three BHK) often a shared dwelling between independent family units which can be indicated by multiple mailboxes and different entrances.
- There may be a gate but usually no high wall present around the house.
- Apartment buildings often have outdoor staircases, may have a gate entrance to building but generally not part of a complex or gated community.
- Often mostly concrete but some have additional materials such as glass/wood/brick, etc.
- Apartments often have private balconies.

Picture 2: HT-4

Category 3 Dwellings (HT-3): Lower Middle-Class Housing

- Apartments and houses are most often made only of concrete.
- Windows are often smaller.
- Houses are small, often two-three rooms with concrete roofs, usually only one level.
- Often in neighbourhoods containing Housing Type 2s and Housing Type 1s.
- Sometimes interspersed with commercial shops/denser neighbourhoods.
- Apartment buildings may often be above small shops, often no gate around apartment building.
- May often have shared balconies across units.
- JNNURM social housing built for slum relocation; these buildings are often green and white with JNNURM printed on the side. Small concrete open windows/no glass, inside staircases, community bathrooms.

Picture 3: Housing type 3

Category 2 Dwellings: Informal slum settlements (HT2)

- One-room pukka row house.
- Corrugated metal roof.
- Densely packed.
- Often not located on a main street, behind buildings, down gullies.

- Often in neighbourhoods containing 2's and 1's, and small one room commercial businesses.
- Few windows, small windows, often shutters not glass.
- One entrance.

Picture 4: Housing type 2

Category 1 Dwellings: Informal shack Settlements (HT1)

- Self-built dwelling often made from: reclaimed wood, fabric, tarpaulin, corrugated metal, sackcloth.
- Often not located on street-fronts, often located in vacant lots, behind buildings, on sidewalk, road medians, small green spaces, large slums, under overpasses, construction sites.
- Can be two floors or one floor.
- Can be a family living inside of a larger vacant and abandoned/under-construction non-self-made structure, but often using self-made materials within that building (tent, etc.).

Picture 5: Housing type 1

Appendix 5: Example Listing Map

The map shows how the buildings were drawn on the map in a spatial manner and allocated a Housing Type also, with a summary given in the margin.