INFERRING COVID-19 CASES FROM SLUMS AND NON-SLUM AREAS IN MUMBAI

MURAD BANAJI*, DECEMBER 2020

Data on how many of the city's recorded COVID-19 cases involve the slum population is not released publicly. On the other hand, from the second week of April, ward-wise case data was released, and we can hope to use this in combination with 2011 census data on the slum/nonslum populations in each ward to infer the proportion of cases coming from the slums in the city.

To this end, we divide the city into two compartments which we'll term C_s (the city's slums) and C_n (the remainder of the city). We also divide the city's 24 wards into three strata termed the slum-light stratum S_l, the slum-medium stratum S_m, and the slum-heavy stratum S_h. (The methods to follow could be applied with more than three stata.)

Ward-wise population estimates at the 2011 census are taken from here (in Marathi). Data and details are given in the Appendix.

Any given subpopulation of the city corresponds to a nonnegative vector (a, b, c) with a members in S_l, b in S_m and c in S_h. The three strata thus define a $3D$ vector space $\mathcal{X}$ consisting of all such stratified population vectors. The slum and nonslum populations of the city themselves define positive vectors $v_s, v_n \in \mathcal{X}$ respectively. $\mathrm{span}\{v_s, v_n\}$ is a 2D subspace of $\mathcal{X}$, which we'll term $\mathcal{X}'$. We can say, roughly, that a given subpopulation of the city, represented as a point $x \in \mathcal{X}$, is "well-modelled" by the two compartment model if x lies close to $\mathcal{X}'$. What this means is that the way this population is divided between strata is consistent with the slum/nonslum make-up of each stratum.

Given a nonnegative and nonzero vector $x \in \mathcal{X}$, we define a vector $x' \in \mathcal{X}'$ as follows:

- We first project x orthogonally onto $\mathcal{X}'$ to get a new vector $x'' \in \mathcal{X}'$.
- We then rescale x'' so that the sum of its components is equal to that of x, and in this way obtain x'. (Note that x, being nonnegative and nonzero, cannot be orthogonal to $\mathcal{X}'$, which includes positive vectors, and so $x'' \neq 0$. The rescaling is thus always possible.)
- We can now resolve x' along v_s and v_n to get $x' = x^s v_s + x^n v_n$. If x^s is negative (i.e., x' fails to lie in the convex hull of v_s and v_n), then we redefine x' to be the unique multiple of v_n and with components summing to those of x, namely $(x_1 + x_2 + x_3)/(v_{n,1} + v_{n,2} + v_{n,3})v_n$; and similarly if x^n is negative we insist that x' is a multiple of v_s with components summing to those of x. (In practice this consideration only occurs when we look at cumulative case data for a few of the very early days where data is available. Note that x^s and x^n cannot both be negative as this would imply that $x - x'$ was a nonnegative and nonzero vector, contradicting the fact that it is orthogonal to $\mathcal{X}'$.)

We refer to any vector x' obtained from a positive vector x as above as the "modelled" version of x.

Let C_i^l refer to the cases in the slum-light stratum on day n obtained from ward-wise data, with C_i^m and C_i^h similarly defined. Cases on day n thus define a case-vector $C_i = (C_i^l, C_i^m, C_i^h) \in \mathcal{X}$. Let C_i' be the modelled version of C_i.

We can see C_i' as the closest approximation to the city's case vector on a given day compatible with the two compartment model, and with the same total cases. Note that a (positive) rescaling of C_i leads to a rescaling of C_i'' and hence C_i.

If we resolve C_i' along v_s and v_n to obtain $C_i' = c_i^s v_s + c_i^n v_n$, then by construction c_i^s and c_i^n are nonnegative and estimate daily per-capita cases from the slums and nonslum areas resectively.

*Middlesex University, London, Department of Design Engineering and Mathematics. m.banaji@mdx.ac.uk.

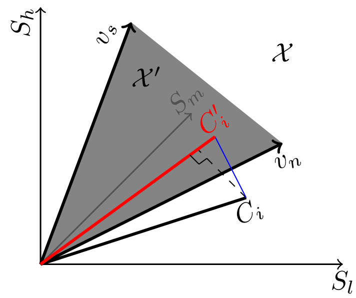

FIG. 1. *Schematic demonstrating the vector C_i and the "modelled version" C_i' on $\mathcal{X}'$.*

In practice, for robustness, it makes most sense to work with weekly averages rather than daily case data.

Appendix A. Mumbai 2011 demographic data. Wards with up to 33% of their population in slums are classified as slum-light. These are wards A, B, C, D, E, H/W, K/W, G/S, R/C and T which together contain about 29% of the city's population. 25% of the population of these wards lives in slums. Wards with 61% or more of their population in slums are classified as slum-heavy. These are wards R/N, K/E, H/E, P/N, L and M/E, which contain about 36% of the city's population. 76% of the populations of these wards lives in slums. The remaining wards (N, M/W, F/S, F/N, S, R/S, G/N and P/S) contain another 36% or so of the city's population, and make up the slum-medium stratum. 51% of the population of these wards lives in the slums.

ward	total	slums	nonslums	% in slums
A	185014	22282	162732	12%
B	127290	12711	114579	10%
C	166161	16571	149590	10%
D	346866	34699	312167	10%
E	393286	124194	269092	32%
G/S	377749	124306	253443	33%
H/W	307581	82552	225029	27%
K/W	748688	215678	533010	29%
R/C	562162	172849	389313	31%
T	341463	85560	255903	25%
F/S	360972	180128	180844	50%
F/N	529034	238128	290906	45%
G/N	599039	361674	237365	60%
P/S	463507	230829	232678	50%
R/S	691229	414395	276834	60%
M/W	411893	164992	246901	40%
N	622853	249229	373624	40%
S	743783	408442	335341	55%
R/N	431368	281151	150217	65%
K/E	823885	572818	251067	70%
H/E	557239	388923	168316	70%
P/N	941366	708247	233119	75%
L	902225	758108	144117	84%
M/E	807720	685994	121726	85%
total	12442373	6534460	5907913	53%

TABLE 1

Total population in Mumbai's wards at the 2011 census from MCGM data (link in text above). The slum-light, slum-medium and slum-heavy strata consisting of 10, 8 and 6 wards, respectively are shown.