

# QUANTIFYING SPATIAL MISALLOCATION IN CENTRALLY PROVIDED PUBLIC GOODS

SIVA ATHREYA AND ROHINI SOMANATHAN

ABSTRACT. We show how an optimization algorithm can be used to approximately quantify the costs to users of spatial misallocation in centrally provided public goods. This method can be employed to evaluate the large programs of public good construction that have been central features of economic plans in many developing countries. We apply these methods to the allocation of post-offices in an administrative block of South India between 1981-1991 and find that more appropriate choices for post office locations could have reduced aggregate costs of travel to citizens in this area by at least 20%.

May 18, 2004

## 1. INTRODUCTION

Many countries, at some point during their development process, have initiated large construction programs to make public goods accessible to their citizens. A major thrust of the Second Five Year Plan in Indonesia in the 1970s was to improve access to primary schools ([4]). At roughly the same time, the Indian government, under the *Minimum Needs Programme* began constructing a variety of infrastructural projects in rural areas ([10]). The stated objective in these and other similar programs was to bring amenities close to their potential users.

There are plausible reasons why factors other than the proximity of users may have influenced actual allocations. The political power of local elites or differences in the strength of collective action across settlements could bias allocations in favor of some

---

We are grateful to the Policy and Planning Research Unit of the Indian Statistical Institute for financial support and to Abhijit Banerjee, Satish Rao and E. Somanathan for very useful discussions.

sections of the population ([14]). A substantial body of empirical work suggests that social cohesion and local collective action does vary across communities and that these variables influence the level and quality of public goods ([1] [11] [7]). This appears to be true even when public goods are centrally financed ([8] [3]). While the existing empirical literature has established the importance of a community's characteristics in determining its share of total allocations, there is no work, to our knowledge, on measuring the total welfare loss resulting from these biased allocations.

In this paper, we propose a method which can be used to quantify the costs of misallocation in public goods, relative to an allocation which minimizes the aggregate distance traveled by users. Specifically, we are concerned with the problem in which at some initial time period, there is an arbitrary spatial distribution of users residing in a large but finite set of locations and an existing set of public facilities. Citizens are identical in all respects other than their location and use the facility closest to them. Their travel costs are linear in the distance traveled to a facility. During a subsequent period, a planner allocates a given number of additional facilities. At the end of the program, we are interested in computing the difference between travel costs corresponding to the observed location of facilities and those that would occur if facilities were located with the objective of minimizing such costs. This difference provides us with a measure of the cost of *misallocating* public goods in the area.

With a small number of locations, the allocation of a set of public facilities which minimizes aggregate travel costs can be easily computed. This is because the number of possible configurations is small and a comparison of the travel costs associated with them is all that is required. As the number of habitations increases, the number of computations required increases exponentially and this optimization problem becomes intractable. Problems of this type have been shown to be NP-hard [9]; in other words, for large numbers of users and locations, the number of calculations needed for an optimal solution using any possible algorithm is so huge that a computer cannot do them in a reasonable amount of time.

A variety of algorithms have been developed to obtain  $\epsilon$ -suboptimal solutions to such problems. That is, suppose  $Z$  is the optimal cost for the above problem,  $Z_\epsilon$  is said to be  $\epsilon$ -suboptimal if  $Z_\epsilon < (1 + \epsilon)Z$ . For this *epsilon* price in optimality, the algorithms are able to provide allocations which can be computed in reasonable (or polynomial) time. Such algorithms guarantee a value for the objective function that does not deviate from the optimized value by more than a  $(1 + \epsilon)$  factor. In this paper, we show that our public goods location problem can be written as a version of a problem in this literature and this allows us to adapt an existing algorithm to compute an allocation for our problem in which travel costs are at most 1.61 times those in the optimal allocation. The difference between computed costs based on the allocation by the algorithm and those based on the observed location of public goods is a lower bound for the cost of misallocating public goods.

We apply the algorithm to data on the location of villages and post offices in an administrative block in South India for two census years, 1981 and 1991. We use post offices as our public amenity since the services they provide are fairly uniform, they are used by most households and do not have any close substitutes in rural India. We focus on a region in South India because geo-coded spatial data is available for this part of the country. We combine data on village locations with census data on village populations and post offices. It would have been ideal to use data from the 1971 Census, since the rapid growth of infrastructural facilities in rural areas began in the 1970s, but we were not able to obtain village level for this period. Between 1981 and 1991, there was a 23% increase in the number of villages with post office facilities in the area we study. Our objective is to find the allocation of these additional post offices which would have minimized aggregate travel costs in 1991.

The difference in aggregate distance traveled under the allocation made by the algorithm and the actual allocation recorded in the census data for 1991 is a little over 21%. Given that costs associated with the computed allocation could be as much as 1.61 times the costs corresponding to the minimized solution, this is a lower bound on the cost savings

that could have been achieved by the planner. Such deviations from optimal allocations could have resulted because the decision makers had a different objective function, incorporating, for example, the political influence enjoyed by different villages or other community characteristics discussed in the literature. Or, they may simply reflect the difficulties of solving the travel cost minimization problem.

Section 2 contains the statement of our problem and a brief description of the algorithm we use to obtain an approximate solution. Section 3 applies the algorithm to data on the location of villages and post offices in the administrative block of Vriddhachalam in South India, and compares travel costs from the resulting allocation with actual travel costs. Section 4 concludes with a mention of some possible directions for future research. A detailed description of the algorithm and the mathematical analysis is presented in Section 5.

## 2. SOLVING THE PLANNER'S PROBLEM

In keeping with the standard notion of public goods, we assume that they have no marginal costs of additional users. A planner observes an existing distribution of these goods, and would like to allocate a fixed number of additional facilities to minimize the total distance traveled by users after the allocation. A precise statement of this problem is given below.

**Problem 2.1.** *Suppose there are  $n$  villages with locations given by  $V = \{v_1, v_2, \dots, v_n\}$  in a specified geographical area and there are  $k_1$  facilities located at  $\{s_1, \dots, s_{k_1}\}$  (a subset of  $V$ ). Let  $P_j$  be the population in village  $j$ . The planner wants to allocate an additional  $k_2$  facilities in a manner which minimizes aggregate distance traveled by the entire population.*

*In particular define*

$$Z(n, S) = \sum_{j=1}^n P_j \min_{1 \leq i \leq k_1+k_2} \|v_j - s_i\|,$$

*where  $S = \{s_{k_1+1}, s_{k_1+2}, \dots, s_{k_1+k_2}\} \subset \mathbb{R}^2$  is the set of positions of the  $k_2$  facilities and  $\|v_j - s_i\|$  denotes the Euclidean distance between  $i$ -th village and the  $j$ -th facility. The*

planner needs to find an allocation  $S$  so as to achieve  $Z$  given by

$$(1) \quad Z = \min_{S:|S|=k_1+k_2} Z(n, S),$$

where  $|S|$  is the number of elements in the set  $S$ .

This problem is a constrained version of the well known  $k$ -median problem, in which, starting with no facilities, a total of  $k = k_1 + k_2$  facilities are allocated to minimize aggregate distance traveled. For large values of  $n$  this problem is known to be NP-hard [9]. The main difference between the  $k$ -median problem and Problem 2.1 is in the location of  $k_1$  facilities. These cannot be changed in Problem 2.1 and the planner allocates only  $k - k_1$  facilities to minimize the same objective function.

A number of available algorithms can be used to obtain  $\epsilon$ -suboptimal solutions to the  $k$ -median problem (see for instance [2] and [13]). For obtaining a solution for Problem 2.1, we adapt the algorithm presented in [6] for a related problem called the facility location problem. This allows us to obtain a value of the objective function in Problem 2.1 which is at most 1.61 times the minimized value.

We briefly describe the algorithm here for the case when  $k_1 = 0$  and show how it can be adapted to provide a solution to Problem 2.1. Section 5 contains details of the algorithm and we show there that the travel costs resulting from our solution to Problem 2.1 are at most 0.61-suboptimal.

The algorithm proceeds by opening facilities at selected villages and connecting villages to these facilities. People in a village will access the facility that the village is connected to. The travel cost for each village is therefore equal to the distance of the facility from the village multiplied by the village population.

It begins with some initial cost of opening a public facility at village  $i$ . Denote this by  $f_i$ . We discuss how to choose  $f_i$  in the next paragraph. Let  $d_{ij}$  denote the distance from village  $i$  to village  $j$ . We start at time 0. At this time, each village  $j$  has a budget,  $B_j$ ,

which is initialized to 0 and a potential facility. The budget of the village increases by 1 in each time period, as long as it is unconnected to an open facility. In each period, if the budget of a village  $j$  is bigger than the distance to a potential facility  $i$ , it offers that facility the difference  $(B_j - d_{ij})$  times its population  $P_j$ . A facility  $i$  is opened the moment that contributions to it reach the facility cost  $f_i$ . Once a facility is opened, all villages with positive contributions to it are connected to it. If a village  $j$  is already connected to some other facility  $i'$ , then its offer to facility  $i$  in each period is equal to  $P_j * \max(d_{i'j} - d_{ij}, 0)$  (the saving in travel cost for  $j$  incurred by switching from  $i'$  to  $i$ ). The algorithm continues until all villages are connected to some open facility and the total number of facilities opened is equal to the desired number  $k$ .

For the case where  $k_1 = 0$ , we set all  $f_i$  to be a constant  $f$ . From the discussion on page 244-245 in [13], or the detailed description of the algorithm in Section 5, it is easy to observe that the number of facilities opened is a decreasing function of facility cost  $f$ . Hence if the number of facilities opened by the algorithm is less (greater) than  $k$ , the facility cost  $f$  is lowered (raised) and the algorithm is run again, until the number of additional facilities, after all connections have been made, is exactly equal to  $k$ . For Problem 2.1, we must ensure that our algorithm opens  $k_1$  facilities at villages that have pre-existing facilities and an additional  $k_2$  facilities elsewhere. For this to happen we set  $f_i = 0$  if village  $i$  has an opened facility in period 1, else we set it to a constant  $f$ . By design the algorithm will open the existing  $k_1$  facilities. If the number of additional facilities opened by the algorithm is less (greater) than  $k_2$ , the facility cost  $f$  is lowered (raised) and the algorithm is run again, until the number of additional facilities, after all connections have been made, is exactly equal to  $k_2$ .

At times a situation may arise where one cannot find a  $f$  that produces  $k_1 + k_2$  facilities. In such cases an approach laid out in pages 247-251 [13] is used. It involves a procedure called randomized rounding. However this has a doubling effect on the optimality factor. For our dataset we were able to find such an  $f$  and did not have to use this (See Section 5).

TABLE 1. Summary of Results

<b>Quantity</b>	<b>1981 Actual</b>	<b>1991 Actual</b>	<b>Algorithm 1991-1981</b>
Total travel cost in kms.	133747	115197	94887
Average population in 1991 of of villages without post offices.	1120	1079	998
Average distance in kms. to nearest post office (for villages without post offices)	1.59	1.51	1.37
Population in 1991 of largest village without post office	3407	2634	2502

### 3. POST OFFICES OF SOUTH INDIA

In this section we apply the algorithm for Problem 2.1 to data from the administrative block of Vriddhachalam in South India. We use data from the Census of India for 1981 and 1991 on village populations and the availability of post offices in each village. We combine these with geo-coded data for village locations available in the South India Population Information System which has been put together by the French Institute in Pondicherry.

According to the census data, out of a total of 156 villages in the area, there were 70 villages with post office facilities in 1981. This figure went up to 86 in 1991.<sup>1</sup> We calculate the travel cost per village as the product of the population of the village and the distance, in kilometers, to the nearest village with a post office facility.<sup>2</sup> We have used the Euclidean distance between the two villages rather than the actual road distance. The data on road distance is hard to obtain. In addition, the shortest route taken to a post office from any village in this area is not likely to be on paved roads, but through fields and footpaths. Since we are dealing with facilities that are fairly numerous in fairly flat terrain, it is reasonable to assume that ordering of distances based on routes actually taken by villagers will not be too different from the Euclidean distances.

Our results are presented in Table 1. For purposes of comparison, we list some characteristics of the allocation in 1981 and 1991 which we observe in the census data. Of the 16 new post office facilities allocated during this period, we have five matches between algorithm and actual allocations. Aggregate travel costs in 1991 are 21% higher than those corresponding to the algorithm for Problem 2.1. As can be seen from Table 1, the cost saving achieved by the algorithm seems to result both from locating post offices in more densely clustered areas as well as in bigger villages.

#### 4. DISCUSSION

We have proposed a method, based on the literature on optimization algorithms, for calculating a lower bound on the extent of misallocation in programs of public good construction. The procedure outlined in this paper allows for allocations to be constrained by the distribution of pre-existing facilities. In this sense it solves a two-period problem. The algorithm used is an adaptation of static algorithm in the literature which has an

---

<sup>1</sup>There were 11 villages for which post office facilities were recorded in the census data of 1981 but they were not marked in the 1991 census data. Conversations with census officials led us to believe these were errors of omission in 1991. Our figure of 86 for 1991 therefore includes these.

<sup>2</sup>This means that the village has either a post office or a post and telegraph office.



optimality factor of no less than 1.61 and we show that our algorithm maintains the same optimality factor.

The two-period nature of our problem deserves emphasis. Had we not taken account of pre-existing facilities, the difference between the observed allocation and the one given by the algorithm could not be attributed to misallocation. On the other hand, had period 2 not been the final period, we would not be able to term the differences between actual and computed allocations as *inefficiency*. The actual allocation may be part of a solution to a dynamic problem in this class, which will not, in general, minimize travel costs every period. It is for these reasons that our methods are best suited to evaluate time-bound programs of expansion in public amenities in which the goals of the program are stated in terms of benefits and costs in the final period alone. An obvious extension of this work would be to examine multi-period problems of public good locations. This is an open area in terms of both analytical solutions for small finite populations and algorithms for larger populations.

Geo-coded data is rapidly becoming available for many parts of the world. The methods proposed here can be used to generate variables which capture the quality of governance. The difference between actual and optimal solutions to the spatial location problem can be correlated with other characteristics of geographical areas to provide insight into the mechanisms used by governments to provide public goods.

## 5. ALGORITHM AND ANALYSIS

In this section will present a method to find an allocation that will provide an 0.61-suboptimal solution to Problem 2.1. The first two steps of the solution involve adapting an algorithm presented in [6] for a closely related problem. This algorithm opens a certain number of facilities and connects villages to them with the objective of minimizing total travel plus facility costs. The last step of the solution adjusts opening costs to so as to

open a pre-assigned number of facilities. In Section 5.1 we provide a detailed analysis of the solution.

### Steps to solve Problem 2.1

Let  $V$  be as before. Let  $d_{ij} = \|v_j - v_i\|$  be the distance between villages.

- (1) For each village  $i$ , we define  $f_i$  to be the cost of opening a facility at village  $i$ . We set  $f_i = 0$  for all  $i$  if the facility  $i$  was opened in period 1. For all other  $i$  set  $f_i = f$  an initial constant.
- (2) In this step we perform the following algorithm for the chosen  $f$ .

#### Algorithm A (From [6]).

- (a) We introduce a notion of time. The algorithm starts at time 0. At this time, all villages are unconnected, all facilities are unopened, and the budget of every village  $j$ , denoted by  $B_j$ , is initialized to 0. At every moment, each village  $j$  offers some money from its budget to each unopened facility  $i$ . The amount of this offer is computed as follows: If  $j$  is unconnected, the offer is equal to  $P_j * \max(B_j - d_{ij}, 0)$  (i.e., if the budget of  $j$  is more than the cost that it has to pay to get connected to  $i$ , it offers to pay this extra budget to  $i$ ); If  $j$  is already connected to some other facility  $i'$ , then its offer to facility  $i$  is equal to  $P_j * \max(d_{i'j} - d_{ij}, 0)$  (i.e., the amount that  $j$  offers to pay  $i$  is equal to the amount  $j$  would save by switching its facility from  $i'$  to  $i$ ).
- (b) While there is an unconnected village, increase the time, and simultaneously, increase the budget of each unconnected village at the same rate (i.e., every unconnected village  $j$  has  $B_j = t$  at time  $t$ ), until one of the following events occur. If multiple events occur at the same time, process them in an arbitrary order.
  - (i) For some unopened facility  $i$ , the total offer that it receives from villages is equal to the cost of opening  $i$ . In this case, we open facility  $i$ , and for

- every village  $j$  (connected or unconnected) which has a non-zero offer to  $i$ , we connect  $j$  to  $i$ .
- (ii) For some unconnected village  $j$ , and some facility  $i$  that is already open, the budget of  $j$  is equal to the connection cost between  $j$  and  $i$ . In this case we connect village  $j$  to facility  $i$ .
- (c) Perform this step after (a) and (b) are completed and all villages are connected to a open facility. Reassign the connections of each village to the nearest open facility.
- (3) The  $f$  chosen in Step 1, may open exactly  $k_1 + k_2$  facilities. If it does then we are done. Otherwise notice that the algorithm will open more facilities for small values of  $f$  and less facilities for large values of  $f$ . Hence via binary search or otherwise arrive at a value of  $f$  that opens  $k_1 + k_2$  facilities

For our dataset the correct value of  $f = 3400$ . It maybe happen that one cannot find a  $f$  that produces  $k_1 + k_2$  facilities then an approach laid out in pages 247-251 [13] could be used. It involves a procedure called randomized rounding. However this has a doubling effect on the optimality factor.

**5.1. Analysis.** In this section we discuss the validity and the optimality of the solution provided above. For Problem 2.1, we need to address two issues. First is that, the solution opens  $k_1$  facilities at places which had facilities in time period 1. This is achieved because once a facility  $i$  has  $f_i = 0$ , the algorithm A will open it straightaway. This is seen easily on a second reading of the algorithm. As soon as we find an  $f$  that opens  $k_1 + k_2$  facilities, we have opened the additional  $k_2$  facilities as desired.

The second issue is of the optimality factor. This analysis relies on the ideas presented in pages 244-245 [13] and the result in [6]. For completeness we present the statement and the argument in its entirety below. We present the broad intuitive idea first. In [6] it is shown that Algorithm A provides a 0.61-suboptimal solution for a closely related problem called the Facility Location problem. It so happens that the objective function and of this problem and Problem 2.1 are very similar. The key difference being that the

former does not impose any direct bound on the number of facilities opened. However knowing that Step 3 was successful in our solution and using a standard theorem in the Linear programming literature called the LP-duality Theorem we are able to obtain the optimality factor of our solution to Problem 2.1.

**Proposition 5.1.** *Assume that Step 3 in our Solution to Problem 2.1 is successful. Then the allocation provided by this solution is a 0.61-suboptimal solution of Problem 2.1.*

Before we provide the proof of this proposition we need some preliminary facts and results from the Linear Programming literature.

5.1.1. *Preliminaries.* Let  $f_i, P_j, d_{ij}$  be as before. We proceed to state the result from [6]. Consider the following integer Linear programming problem called the Facility Location problem (FLP) and the dual problem associated with the relaxed LP<sup>3</sup>.

FLP	Dual
(2) minimise $\sum_{i,j=1}^n x_{ij}P_jd_{ij} + \sum_{i=1}^n y_i f_i$ , Subject to: (3) $\sum_{i=1}^n x_{ij} \geq 1 \quad \forall j$ . (4) $y_i - x_{ij} \geq 0 \quad \forall i, j$ . (5) $y_i \in \{0, 1\}, x_{ij} \in \{0, 1\} \quad \forall i, j$ .	(6) maximise $\sum_{j=1}^n \alpha_j$ , Subject to: $\alpha_j - \beta_{ij} \leq P_j d_{ij}, \quad \forall i, j$ . $\sum_{i=1}^n \beta_{ij} \leq f_i \quad \forall j$ . $\alpha_j \geq 0, \beta_{ij} \geq 0 \quad \forall i, j$ .

The above problem is known as the Facility Location problem. The variable  $x_{ij} = 1$  implies that village  $j$  is connected to facility in village  $i$  (i.e. people from village  $j$  would be using the facility in village  $i$ ) and  $y_i = 1$  implies that the facility in village  $i$  is opened. (3)

<sup>3</sup>The relaxed problem is the facility location problem but with constraint (5) replaced by  $x_{ij} \geq 0$  and  $y_i \geq 0$ .

ensures that every village is connected to at least one facility and (4) ensures that facility is open. The objective function calculates the net costs of opening facilities and traveling to them.

**Theorem 5.1.** (From [6]). *Algorithm A provides a 0.61-suboptimal solution to the facility location problem (2). Moreover the allocation provided by Algorithm A satisfies the*

$$(7) \quad \sum_{i,j=1}^n x_{ij} P_j d_{ij} + (1.61) \sum_{i=1}^n y_i f_i \leq (1.61) \sum_{j=1}^n \alpha_j,$$

where  $\alpha_j$  is a feasible solution of the dual problem (6).

Next we observe that Problem 2.1 can be represented as a integer linear programming model as follows. Rename the villages by the index set  $I = \{1, \dots, n\}$ . Let  $I_1 \subset \{1, \dots, n\}$  be the set of villages that have opened facilities in period 1. We now present the LP and the dual to the relaxed LP.

Integer LP ( $\equiv$ Problem 2.1)	Dual
<p>(8) minimise <math>\sum_{i,j=1}^n x_{ij} P_j d_{ij}</math></p> <p>Subject to:</p> $\sum_{i=1}^n x_{ij} \geq 1 \quad \forall j.$ $y_i - x_{ij} \geq 0 \quad \forall i, j.$ $\sum_{i \in I \setminus I_1} -y_i \geq -k_2$ $\sum_{i \in I_1} y_i \geq k_1$ $y_i \in \{0, 1\},$ $x_{ij} \in \{0, 1\} \quad \forall i, j.$	<p>(9) maximise <math>\sum_{j=1}^n \alpha_j - ak_2 + bk_1,</math></p> <p>Subject to:</p> $\alpha_j - \beta_{ij} \leq P_j d_{ij}, \quad \forall i, j.$ $\sum_{j=1}^n \beta_{ij} \leq -b \quad \forall i \in I_1.$ $\sum_{j=1}^n \beta_{ij} \leq a \quad \forall i \in I \setminus I_1.$ $\alpha_j \geq 0, \beta_{ij} \geq 0 \quad \forall i, j.$ $a, b \geq 0$

Note that in the LP the constraints force  $y_i = 1 \forall i \in I_1$  and in the dual problem the constraints force  $b = 0$  and  $\beta_{ij} = 0$  for all  $i \in I_1, j \in I$ .

5.1.2. *Proof of Proposition 5.1.* By our hypothesis we have found  $f_i$  (equivalently a value of  $f$ ) that opens  $k_1 + k_2$  facilities in our solution for Problem 2.1. Now consider the allocation  $(\bar{x}, \bar{y})$  and  $(\bar{\alpha}, \bar{\beta})$ <sup>4</sup> given by our algorithm. Note that these are feasible solutions to (9) and (8) respectively. From (7), we have that

$$(10) \quad \begin{aligned} \sum_{i,j=1}^n \mathbf{x}_{ij} P_j d_{ij} &\leq 1.61 [\sum_{j=1}^n \bar{\alpha}_j - \sum_{i=1}^n f_i \mathbf{y}_i] \\ &= 1.61 [\sum_{j=1}^n \bar{\alpha}_j - f k_2] \end{aligned}$$

Suppose  $(\hat{x}, \hat{y})$  and  $\hat{\alpha}, \hat{a}$  are optimal solutions of the relaxed problem (8) and (9) then

$$(11) \quad Z \equiv \sum_{i,j=1}^n \hat{x}_{ij} P_j d_{ij} = \sum_{j=1}^n \hat{\alpha}_j - \hat{a} k_2$$

where  $Z$  is as in (1). The first equivalence is straightforward by definition. The second equality is by the LP-duality theorem (see page 95 and 244 [13]). Using (10) and (11) we have that the algorithm has found a 0.61-suboptimal solution for Problem 2.1.  $\square$

**Remark:** The above proposition does have an analogue if we are not able to find an  $f$  that opened exactly  $k_1 + k_2$  facilities. The optimality factor analysis is more complicated. The interested reader can imitate the arguments in pages 245-248 to achieve a 2.32-suboptimal solution at all times for Problem 2.1.

---

<sup>4</sup>By the comment on the variables it is easy to see how to find  $\bar{x}, \bar{y}$  given an allocation. From Algorithm A, one observes that  $\bar{\alpha}_j$  is  $P_j B_j$  at the time of connection and  $\bar{\beta}_{ij}$  is the net contribution given by village  $j$  to village  $i$ .

## REFERENCES

- [1] Alberto Alesina, Reza Baqir and William Easterly, *Public Goods and Ethnic Divisions*, Quarterly-Journal-of-Economics, 114(4), pages 1243-84, 1999.
- [2] S. Arora, P. Raghavan, and S. Rao, *Approximation schemes for Euclidean  $k$ -medians and related problems* STOC '98 (Dallas, TX), pages 106–113, ACM, New York, 1999.
- [3] Abhijit Banerjee and Rohini Somanathan, *The political economy of public goods: Some evidence from India*, mimeo, MIT and University of Michigan, 2004.
- [4] Esther Duflo, *Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an unusual policy experiment*, American Economic Review, 91(4), pages 795-813, 1991.
- [5] M.L. Fisher and D.S. Hochbaum, *Probabilistic analysis of the planar  $k$ -median problem* Mathematics of Operation, (5), No. 1, pages 27-35, 1980.
- [6] K. Jain, M. Mahdian, E. Markakis, A. Saberi, and V. V. Vazirani, *Greedy facility location algorithms analyzed using dual fitting with factor revealing LP* Approximation, randomization, and combinatorial optimization (Berkeley, CA, 2001), 127–137, Lecture Notes in Comput. Sci., 2129, Springer, Berlin, 2001.
- [7] Asim Ijaz Khwaja, *Can good projects succeed in bad communities*, mimeo, Harvard University, 2001.
- [8] Brian Knight, *Parochial interests and the centralized provision of local public goods: evidence from congressional voting on transportation projects*, Journal of Public Economics, 88(3-4), pages 845-66.
- [9] N. Megiddo and K. Supowit, *On the complexity of some common geometric location problems* SIAM journal of computation, (13), No.1, pages 183-196, 1984.
- [10] Planning Commission, Government of India, *Fifth five year plan, 1974-1979*, Controller of Publications, Delhi, 1976.
- [11] Robert D. Putnam, *Making democracy work: civic traditions in modern Italy*, Princeton University Press, 1993.
- [12] J.H. Lin and J.S. Vitter, *Approximation Algorithms for geometric median problems* Information Processing letters, (44), pages 245-249, 1992.

- [13] V. Vazirani, *Approximation Algorithms* Springer-Verlag, New York, 1999.
- [14] Barry R. Weingast, Kenneth A. Shepsle and Christophe Johnsen, *The political economy of benefits and costs: a neoclassical approach to distributive politics*, *Journal of Political Economy*, 89(4), pages 642-64.

Siva Athreya, Indian Statistical Institute New Delhi. Email: athreya@isid.ac.in

Rohini Somanathan, Department of Economics, University of Michigan and Indian Statistical Institute, New Delhi. Email: rohini@umich.edu