

# QUANTIFYING SPATIAL MISALLOCATION IN CENTRALLY PROVIDED PUBLIC GOODS.

SIVA ATHREYA AND ROHINI SOMANATHAN

ABSTRACT. We use an optimization algorithm to solve a two-period planner's problem of spatially allocating public goods. We apply the algorithm to data on the location of post-offices in South India between 1981-1991 and show that more appropriate choices could have reduced travel costs by at least 20%.

**JEL Classification Codes: C63, H41, D78.**

March 13, 2007

## 1. INTRODUCTION

Many countries, at some point during their development process, have initiated large construction programs aimed at creating universal access to public amenities such as schools, postal and communication services, hospitals and health centers. There is now substantial evidence within the social sciences on spatial misallocation in such spending. This could result from the uneven political leverage of individuals and social groups, agency problems in implementing desired allocations, bureaucratic incompetence, or simply a lack of information about optimal allocations ( [1] [2] [4] ). Our focus here is on quantifying the total welfare loss associated with these multiple factors. We propose a simple algorithm, adapted from the recent computer science literature, which can be used to approximate the optimal allocation of a fixed number of facilities given a spatial distribution of users and existing facilities. These methods can be useful in evaluating programs of public good expansion and can also serve as planning tools that guide future spending.

We consider a setting in which, in some initial period, the population is distributed across a large but finite set of locations. Some of these locations already have the public goods in question. It is typically infeasible, for either political or practical reasons, to move facilities once they have been constructed and we're therefore interested in the problem of optimally allocating additional facilities given this initial distribution. We assume that citizens are identical in all respects other than their residence and that social welfare is decreasing linearly in the aggregate distance travelled by the population. If a solution to this two-period problem were available, we could compute travel costs corresponding to

---

We are grateful to the Policy and Planning Research Unit of the Indian Statistical Institute for financial support and to Jim Adams, Abhijit Banerjee, Emily Gray, Satish Rao, Gary Saxonhouse and E. Somanathan for useful discussions.

this allocation. To evaluate a past program of public good construction, we could use the difference between this computed value and the actual distance travelled as a measure of *misallocating* public goods in the area. Our purpose here is to arrive at such a measure.

With a small number of locations, computing the optimal allocation is straightforward. 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 [5]; 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 data 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. We use our algorithm to obtain an allocation of these additional post offices which would have (approximately) minimized aggregate travel costs in 1991.

We find that aggregate travel costs corresponding to the observed allocation of post offices in 1991 are 21% higher than the costs associated with the allocation suggested by our algorithm. Since the algorithm is constrained to allocate post-office to villages that had post-offices in 1981, it gives us a measure of the costs of misallocation that occurred between 1981 and 1991. Also, 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.

We proceed in the next section with a formal statement of our problem and a 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 Section 4 discusses of some of the limitations of the methods presented here and speculates on how these can be overcome in future research.

## 2. 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  $S$ .

### Algorithm:

We adapt the algorithm presented in [3] for solving a closely related problem, commonly referred to as the *facility location problem*. The facility location problem is one in which there are no facilities initially, and each new facility is associated with a cost. New facilities are opened to minimize the sum of facility costs and aggregate travel costs. We adapt the algorithm for this static problem to a two-period problem in which there are a fixed number of open facilities in the initial period and the total number opened subsequently

correspond to what is actually observed in the data. We are able to show that the use of this algorithm results in a value of the objective function in Problem 2.1 which is at most 1.61 times the minimized value. The algorithm is described below. A formal proof of its optimality factor is available from the authors.

**Step 1:** Here we assume  $k_1 = 0$ . 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.

Let the cost of opening a public facility in village  $i$  be  $f_i$ . 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  $k$ .

It turns out that the number of facilities opened is a decreasing function of facility costs  $\{f_i\}$  (see the discussion on page 244-245 in [6]). Hence if the number of facilities opened by the algorithm is less (greater) than  $k_2$ , the facility costs are lowered (raised) and the algorithm is run again, until the number of additional facilities, after all connections have been made, is exactly equal to  $k_2$ .

**Step 2:** Suppose  $k_1 > 0$ . 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 the facility cost  $f_i = 0$  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 [6] 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 an  $f$  that opened the required number of facilities.

Assume that  $k_1 = 0$  and we are able to find a value  $f$  that opens  $k_2$  facilities. In such an event, the main result in [3] says that the above algorithm provides a 0.61-suboptimal solution to Problem 2.1. We note that in the case  $k_1 > 0$ , the modification to the algorithm that we suggest in step 2 is only in choice of the  $f$  and consequently, as long as we are able to find an  $f$  that opens  $k_1 + k_2$  facilities, the results in [3] hold.

### 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 created 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

---

<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.

TABLE 1. Summary of Results

Quantity	1981 Actual	1991 Actual	Algorithm 1991-1981
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

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

The two-period nature of our problem deserves emphasis. By incorporating pre-existing facilities, the algorithm allows us to assign the misallocation to a well-defined period. This was the decade of the eighties in the case of our application. Used on large national data sets, this can potentially allow researchers to correlate such inefficiencies with changes in political and other institutions. On the other hand, our framework has important limitations. We assume that travel costs are linear and that period 2 is the final period—otherwise we would not be able to term the differences between actual and computed allocations as *misallocation*. The actual allocation at any point in time 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 and allow

for non-linear cost functions. These are open areas 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 and methods which offer solutions to spatial optimization problems are potentially valuable in improving the effectiveness with which public goods are provided and in helping us identify lapses in their delivery.

#### 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] Abhijit Banerjee and Rohini Somanathan, *The political economy of public goods: Some evidence from India*, Journal of Development Economics, volume 82, number 2, pages 287-314, 2007.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] V. Vazirani, *Approximation Algorithms* Springer-Verlag, New York, 1999.

Siva Athreya, Indian Statistical Institute, 8th Mile Mysore Road, Bangalore 560059.  
Email: athreya@isibang.ac.in

Rohini Somanathan, Delhi School of Economics. Email: rohini@econ.dse.org