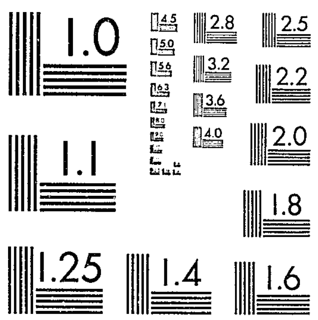


T-88

LOCAL GOVERNMENT SERVICE DELIVERY STRUCTURES: SOME EFFECTS ON PUBLIC BUREAU SUPPLY

by

Roger B. Parks
Elinor Ostrom
Workshop in Political Theory and Policy Analysis
Indiana University

Prepared for delivery at the 1980 meetings of the Association for Public Policy Analysis and Management, Boston, Massachusetts, October 23-25, 1980.

The authors gratefully acknowledge support for the analysis presented in this paper from the National Science Foundation in the form of Grant NSF SES 79-13397; and for the data collection from Grant NSF GI 43949. The f...ngs and opinions expressed herein, however, are solely those of the ...s and do not necessarily reflect those of the Foundation.

...thors appreciate the helpful comments of Debra Dean and Vincent Ostrom ...lier drafts of this paper. The authors also appreciate the high caliber ...fessional assistance we have received from the staff of the Workshop in ...cal Theory and Policy Analysis.

8626

...OP
...L THEORY
...NALYSIS

Indiana University
814 East Third
Bloomington, IN 47405

LOCAL GOVERNMENT SERVICE DELIVERY STRUCTURES:
SOME EFFECTS ON PUBLIC BUREAU SUPPLY

by

Roger B. Parks and Elinor Ostrom

Most conventional analyses of public service delivery employ a unitary model of local governments. In such models, the "government" aggregates consumer preferences, procures and organizes means of service production, and delivers services as a monopoly supplier to constituents. Decisions about output and expenditure levels are assumed to be made by simple referenda or by omniscient and benevolent administrators. But local government service delivery structures are usually not so simple.

Since the early 1960s, scholars have argued for more complex models of public service delivery (e.g., Ostrom, Tiebout, and Warren, 1961; Margolis, 1964). Noting that the local public sector is most frequently composed of several layers of enterprises engaging in a wide variety of exchanges, they argued the need to consider the structure of intra- and interjurisdictional arrangements as influences on service delivery. Margolis, for example, argued that the structure of interorganizational arrangements might make it possible to deal with problems that are less amenable to solution at the level of individual organizations or jurisdictions.

> A consideration of the structure of governments gives a new perspective to old questions. We might ask whether some of the insoluble problems posed in the theory of public expenditures are worked out through the behavior of the structure. That is, does the structure have some of the characteristics of an industry and market, so that there is an interaction among governments which leads to desirable results (Margolis, 1964: 236).

In addition to his concern over the neglect of interorganizational structure, Margolis also criticized analysts of public finance for excessively collapsing

the internal organization of governmental units. Instead of direct democracy or pure hierarchy, most governmental structures are far more complex. As Margolis recognized, these governmental structures may give rise to opportunities for private gain.

> Just as the market can be rigged, the government can be manipulated to protect private interests of some constituents. Just as promoters can orient and stimulate the market, there is a government bureaucracy which can gain from government activities (Margolis, 1964: 236-237).

Despite the cogency of Margolis' argument and those of others (e.g., McKean, 1964), few analysts of local service outputs and expenditures have taken into account overtly the ways the structure of intra- and interorganizational arrangements may affect the performance of local public sector economies. In order to examine the question of how internal and external decision-making structures affect performance, we draw on recent theoretical developments related to theories of public bureau behavior and how interorganizational arrangements affect public bureau behavior.

William Niskanen (1971) was the first to present a rigorous and well-developed formal model of bureau behavior. Mique and Belanger (1974) criticized the Niskanen model and developed an alternative model. Orzechowski (1977) provides a cogent review of the difference between Niskanen and Mique and Belanger, and develops a third model drawing on the theoretical work of Williamson (1964) and the empirical studies by DeAlessi (1969) and Parkinson (1957). Orzechowski predicts that public bureaus, facing demands similar to private firms, will produce output at higher per unit costs and a higher labor to capital ratio. Orzechowski reviews several empirical studies including his own which compare the unit costs and the relative labor to capital ratios for public versus private producers. The consistent finding across a number of different studies is that unit costs of similar output are higher in public than in private firms. Public firms also employ a larger proportion of staff to capital than do private firms (see citations in Orzechowski, 1977; see also Savas, 1978; Pondy, 1968; Lee, 1972).

Work on a theory of interorganizational influences on public bureau behavior has progressed more slowly. Ostrom, Tiebout, and Warren (1961). and Ostrom and Ostrom (1965) argued for the utility of conceptualizing public service delivery structures as "industries." Public service industries, they claimed, might be analyzed using many of the same tools as those employed by economists of the industrial organization persuasion (e.g., Bain, 1959). Consideration of service delivery structures in terms of their monopoly, duopoly, oligopoly, or competitive forms might enable behavioral predictions analogous to those made for private firms in market structures. In an early application of industrial organization concepts to the public sector, Bain, Caves, and Margolis studied the water industry in northern California (1968). But little other empirical or theoretical application of industrial organization concepts occurred until the middle 1970s, in part we believe because of a lack of conceptual tools for characterizing the structure of service delivery arrangements in the public sector.

As a result of National Science Foundation supported studies of the organization of service delivery in metropolitan areas, two similar conceptualizations of service delivery arrangements in the public sector were developed (Ostrom, Parks, and Whitaker, 1974; 1978; Savas, 1978). In both conceptualizations, service delivery arrangements are disaggregated by specific type of service (e.g., general area police patrol, investigation of residential burglaries, radio communications, garbage collection, dry trash collection, newspaper recycling). The participants in the service delivery arrangements are separately classified as producers of the service,

as consumers of the service, or as providers or collective decision-making units that link producer and consumer.[1] Once these three types of participants are separated conceptually, they can be identified empirically for any given service in a particular geographic area (e.g., a city, a county, an SMSA). Matrices can be constructed arraying, for example, all of the producers against all of the consumers (or all groups of consumers for services with attributes of public goods). Each cell in the matrix identifies whether a service link exists between a particular producer and a particular consumer (or group) and, if so, the nature of that service link. Matrices can also be constructed for producer and provider linkages, for provider and consumer linkages, and for linkages between producers of one service and producers of other services that are necessary or useful to the former producers. These service structure matrices, together with computations based upon their sizes and the patterns and types of entries, can then be used to characterize the structure of service delivery arrangements for each service of interest in many different geographic areas (see Ostrom, Parks, and Whitaker, 1978).

By analyzing the relationships between service delivery structures and the behavior of participants within structures of very different forms, we hope to improve our understanding of interorganizational influences on public bureau behavior. Does a public bureau that occupies a monopoly supply position with respect to a large population and across several different services, behave differently than a set of smaller monopolists serving an equivalent total population or a mixed set of more specialized producers of particular services that, in the aggregate, supply an equivalent population? Does the availability of service supply to a given consumer (or group) from two or more different producers lead to inefficient duplication as some would argue, or does the presence of potential competition, even if highly oligopolistic, lead to more vigorous supply efforts by all producers?

In the remainder of this paper we develop several of the components we believe necessary to an understanding of intra- and interorganizational influences on public service delivery. We begin by considering an output or production function for a bureau and discuss how such a function, once empirically estimated, could be used to examine bureau behavior. We then examine the technical exigencies of production functions, and consider the psychological factors likely to influence managerial decision making in public bureaus. In a third theoretical section we discuss how different service delivery structures may affect bureaucratic supply. That is, how might the technical, psychological, and behavioral aspects of a bureaucratic supply model change from one form of service delivery structure to another? Then, in an empirical section, we develop production function estimates for some common police outputs.

Bureau Production Functions

Bureaus, like firms in market situations, are involved in the conversion of inputs into one or more outputs. A number of interesting questions can be raised with respect to the conversion processes adopted by bureaus. Are bureaus technically efficient -- do they obtain the maximum amount of output obtainable with a particular set of inputs? Are bureaus allocationally or price efficient -- do they purchase the least cost combinations of inputs sufficient to supply a given level of output? Are bureaus scale efficient -- do bureaus produce sufficient output to exhaust any scale economies in the

conversion process yet avoid scale diseconomies (if any) associated with further increases in production? Finally, are bureaus efficient in a social welfare sense -- do they tend to produce the level and mix of outputs preferred by the community or communities for whom their supply is intended? (These questions are raised with respect to educational bureaus by Levin, 1976. See also Leibenstein, 1976; and Moran, 1977.)

We believe that the institutional structure within and among bureaus and linking bureaus to consumer communities will have significant effects on answers to these questions. In the private sector pressures on entrepreneurs and managers resulting from the presence of many competitors are believed to lead those decision makers to choose technically and allocationally efficient production processes and input mixes. The necessity of producing outputs that consumers are willing to purchase is thought to push the levels and mixes of outputs produced toward socially efficient frontiers. We think it important to determine whether analogous pressures on decision makers can be found in public sector bureaus and, if so, how pressures for efficiency in each sense of that term used above may vary from one form of institutional structure to another. We begin this investigation by developing a bureau production function. The specific example to be used is that of bureaus supplying police services.

Police bureaus are clearly multioutput producers. Wilson's (1968) simple trichotomy of "law enforcement, order maintenance, and service" illustrates this, as do more elaborate typologies such as Goldstein's (1977) eightfold list of police "objectives." In fact, three- or eightfold typologies cannot do justice to the variety of outputs that are produced by police bureaus. We will be forced to make drastic simplifications when we turn to estimating production functions for police bureaus. Initially, however, we write these functions at a very general level.

We can state an implicit police bureau production function as:

$$F(Y,X,S) = 0 \qquad (1)$$

where Y is a vector of police service outputs, X is a vector of police service inputs, and S is a vector of service condition variables. For each output, Y_i, we can write:

$$Y_i = F_i(Y_i^*, X, S) \qquad (2)$$

where Y_i^* is a vector of all outputs other than Y_i itself. This formulation explicitly acknowledges that police outputs are jointly produced, one with another. Thus, for example, arrests for automobile theft may be produced simultaneously with a reduction in the number of automobile accidents resulting from excessive speed, both by a policy of having officers stop all speeders and request a license check on all stopped vehicles.

While the variety of police bureau outputs is very large, the variety of inputs used by most police bureaus is much less. Police bureaus are highly labor-intensive. Inputs of sworn police officers and civilian employees typically exhaust 80 to 95 percent of police agency budgets. An additional chunk of resources is devoted to the purchase (or lease) of vehicles. For purposes of specifying a production function (and, for later investigation of bureaucratic preferences), we divide sworn police officer inputs into those officers assigned to patrol duties in police agencies and those assigned to other, nonpatrol duties. With this division, we may state the production function for the ith police output as:

$$O_i = K_i P^{a_i} NP^{b_i} C^{c_i} A^{d_i} \left(\prod_{j=1}^{n} S_j^{g_{ij}}\right) \left(\prod_{\substack{j=1 \\ j \neq i}}^{m} O_j^{h_{ij}}\right) \qquad (3)$$

where,

O_i = The quantity of output of the ith type produced,

K_i = A scale factor appropriate for output i,

P = The number of sworn police officers assigned to the bureau's patrol division,

NP = The number of sworn police officers assigned to nonpatrol duties in the bureau,

C = The number of civilians employed by the bureau,

A = The number of automobiles used by the department,

S_j = Service conditions affecting output of the ith type, and

Y_j = All other outputs of the bureau whose production may affect the quantity of output i that is produced.

We have chosen a Cobb-Douglas form for the initial specification of these production functions.[2] If we hold service conditions and all other outputs constant, the coefficients a_i, b_i, c_i, and d_i have ready interpretations. They are the elasticities of output i with respect to their respective inputs. That is, a 1 percent increase in the number of sworn officers assigned to a department's patrol division, P, is predicted to result in a_i percent increase in the output, O_i. The marginal physical product of a particular input can be stated as the ratio of the amount of output being produced to the amount of that input employed, multiplied by the output elasticity of the input. Thus, for officers assigned to patrol, the marginal physical product can be stated as:

$$MPP = a_i \frac{O_i}{P} . \qquad (4)$$

By estimating output functions of this form for various police outputs, we can begin to examine the technical, allocational, and scale efficiency of police bureaus.

For a particular police bureau, we can enter its values for P, NP, C, and A, together with the coefficients for the estimated production function, and determine immediately whether the bureau is doing as well as predicted, or is producing more or less of a given output than predicted by the function (with other outputs held constant). Thus, we can discuss the technical efficiency of particular bureaus. If we know the ratio of wage rates (and automobile costs) confronting a particular police bureau, we can determine whether it has chosen a least cost mix of inputs to produce a given output. The total cost function for a particular choice of inputs is:

$$TC = w_P P + w_{NP} NP + w_C C + r_A A + \Sigma_i r_{0_i} 0_i + FC, \text{ where} \quad (5)$$

w_j = Median wage plus wage related benefits for personnel input j,

r_A = Average cost of owning and operating a police vehicle,

r_{0_i} = Other variable nonwage and nonvehicle expense per unit of output type i, and

FC = Fixed costs of operating a department.

If a given department is operating at the minimum total cost to produce a given mix of outputs, it should employ inputs such that the ratio of their marginal physical products (across all outputs) is just equal to the ratio of their respective wage rates. For officers assigned to patrol and officers assigned to other duties in a department, this requires that:

$$\frac{MPP_P}{MPP_{NP}} = \frac{a_i 0_i \ / \ P}{b_i 0_i \ / \ NP} = \frac{w_P}{w_{NP}}, \quad (6)$$

or that the ratio of patrol officers to officers with other assignments in the department be:

$$\frac{P}{NP} = \frac{a_i w_{NP}}{b_i w_P}. \quad (7)$$

Police departments that employ inputs in ratios other than these may still be operating in a technically efficient manner, but they will not have chosen the least cost combination of inputs to produce a given output.

By examining the estimated output elasticities, we can explore questions of scale efficiencies (holding other outputs constant). All of these elasticities are assumed to lie between zero and one. Whether the sum of the output elasticities, $a_i + b_i + c_i + d_i$, is less than, equal to, or greater than 1.0, indicates whether production is subject to decreasing, constant, or increasing returns to scale of output. With that in mind, we can define a scale efficiency parameter, e_i, as:

$$e_i = (a_i + b_i + c_i + d_i) - 1 \quad (8)$$

Where this parameter is positive, increasing the scale of output would be warranted on efficiency grounds. Where it is negative, decreasing the scale of operation would be in order.

Unfortunately, the factors contributing to scale efficiency are likely to be complex. Economists since at least Adam Smith have thought that economies-of-scale would result from the opportunities for the division of labor and corresponding specialization and mechanization. If such economies are found, then an X percent increase in all inputs should result in a greater than X percent increase in output, and the firm or bureau should expand its output accordingly. Most engineering and process estimates of production technologies indicate that returns to scale are positive over a very wide range (Walters, 1963). Thus, unless some other factor intervenes, one would expect firms and bureaus to be ever-increasing in scale.

Engineering estimates generally do not include the most commonly discussed limiting factor, however. This factor, control loss, is a managerial factor, not subject to easy engineering specification. Control loss results from the dynamics of the flow of information and control among members of the producing organization, particularly across levels of hierarchy within organizations. The information on actual operating conditions that flows upward to managers of such organizations can become distorted as lower level personnel choose to report only those data that place their own activities in a favorable light (Tullock, 1965; Downs, 1967; Williamson, 1967). Directions and commands issued by top administrators tend to become distorted and/or treated as irrelevant to actual conditions as they flow down through hierarchical levels (ibid). Thus, control loss phenomena, which increase with an increase in the size of an organization, may serve as the limiting factor on scale economies (Coase, 1937; Robinson, 1958). The interplay, then, of technical factors contributing to economies of larger scale and of managerial factors that lead to diseconomies of larger scale will determine in specific instances whether the scale efficiency parameter, e_i, is found to be positive, negative, or zero.

A Model of Bureau Supply

In his formal model of bureau behavior, Niskanen (1971: 15) defines bureaus as those organizations that have both of the following characteristics:

(1) The owners and employees of these organizations do not appropriate any part of the difference between revenues and costs as personal income.

(2) Some part of the recurring revenues of the organization derives from other than the sale of output as a per-unit rate.

Most public service delivery agencies meet both these defining characteristics in that neither managers nor employees can legally appropriate as personal income any difference (if one exists) between the lump sum revenues that are the major part of their budget and the costs of operating a department. Considerable disagreement exists about the objective function of bureau managers, but the first condition makes it unlikely that managers

are motivated to reduce operating expenditures below the revenue received by a department. Most public bureaus are observed to spend as much money as they receive in appropriations. Little incentive exists to reduce expenditure levels below approved budget levels as most rewards -- status, increased salary, and other perquisites, for example -- appear to be associated with larger bureau budgets. One way that bureaucrats can increase the probability of larger future budgets is to ensure that all revenue allocated to them in a defined period is spent before the period is over. This line of reasoning leads quickly to assumptions of budget maximizing and/or output maximizing bureaus and bureaucrats. While much contemporary observation of bureaucratic behavior seems consistent with such assumptions, we prefer to delve a bit deeper into motivational concerns before stating any single bureaucratic objective.

The Bureaucrat's Maximand

Unless bureau chiefs are totally under the control of providers (or consumers), at least a part of the determination of bureau output and expenditures should be responsive to their preferences. To explore the influence of bureaucratic preferences, we must consider the objective function of a bureau chief. In particular, what are the likely entries in such a function?

A brief digression on private sector enterprises is useful here. In the classical theory of firms operating in competitive markets, entrepreneur-owners are assumed to have profit maximization as their sole objective. This is plausible in light of their position as claimant to all residual profits in such a firm. This motivation is not necessary to the assumption, however. The entrepreneur in the situationally determinate environment of the competitive market (Latsis, 1972) either chooses profit maximization or market operations force his firm out of business. As Alchian (1950) has demonstrated, it does not matter whether profit maximization is selected consciously or randomly. The result is the same. Only firms that follow profit maximization strategies survive. Not only is the entrepreneur forced to adopt profit maximization, but the analyst of competitive markets is also forced to assume that entrepreneurs in a competitive market are profit maximizers.

As soon as the rigorous conditions of a competitive market are relaxed, neither the entrepreneur nor the manager of a firm is forced to adopt profit maximization. Nor is the analyst forced to adopt this assumption. The theory of the management of the firm has advanced by adopting the utility maximization assumption underlying explanations of consumer behavior, rather than simple profit maximization (Marris, 1964; Williamson, 1964; Alchian, 1965). In the managerial discretion approach to private firm behavior, the manager is assumed to pursue a mixed strategy including striving for a minimally acceptable level of profits for shareholders. Once a minimum level of profits is achieved, theorists predict the manager will trade-off some of any further increments in profits for expenditures that increase the manager's status, income, and leisure. The extent to which a manager is capable of diverting resources for personal rewards "depends on the costs to the stockholders of detecting and policing the managers behavior and effectiveness, i.e., on the costs of enforcing contracts" (Alchian, 1965: 34). These costs vary systematically with the type of ownership arrangements.

It seems reasonable to anticipate that <u>public sector</u> managers might also wish to divert some resources for personal rewards. In spite of possible training and socialization with respect to "the public interest,"

public actors should not be thought of as a race apart from normal human beings (McKean, 1964). Utility maximization assumptions may be as useful, perhaps more so, in predicting public sector behavior as they have been in analyses of private sector managerial behavior.

Utility maximizing private sector actors are assumed to have objective functions that include profits as well as perquisites. That is, some level of profits are necessary to the successful enterprise, with both pecuniary and nonpecuniary rewards to managers determined in part by the level of profits. Profits, too, are typically used as a measure of performance in the private sector. Firms that are more profitable than others facing comparable markets are judged more effective, and their managements are likely to receive higher rewards, both in direct salaries and in other benefits (Lewellen and Huntsman, 1970). To develop a public sector analogy, we must define the public sector equivalent of profits.

Possible substitutes include the concept of a fiscal residuum (Orzechowski, 1977) or that of a bureau's discretionary budget (Niskanen, 1975). Orzechowski defines fiscal residuum as "the difference between tax dollars collected for a public service and the minimum costs of producing that service" (1977: 231). The bureau's discretionary budget is the difference between the maximum budget that the provider would approve for a given output and the minimum total cost of supplying that quantity of output.[3] These concepts both include some attention to the net benefits supplied to a provider's constituents. That is, we assume that no provider would approve a budget and output combination where net benefits to at least a majority of consumers are not greater than (or just equal to) zero. Otherwise, an alternative slate of candidates for provider positions could replace the current provider at the next election (see MacKay and Weaver, 1978, and Langbein, 1980, for explicit statements of this position). Using the concepts of fiscal residuum or discretionary budget, attention immediately turns to discovering whether the bureau is able to capture this surplus through overproduction or through inefficient (higher unit cost) production of a given quantity.

We prefer a different formulation, one which explicitly puts the benefits to consumers into the model. Consequently, we define a performance measure called a benefits residuum, which is the difference between the total value of a bureau's output to citizens of a providing organization and the total cost of producing that quantity of output. That is,

$$BR = \sum_i v_i O_i - TC, \text{ where} \quad (9)$$

BR = Benefits residuum,

v_i = The average per unit valuation of output i across citizens of the providing organization, and

TC = Total cost of producing the sum of the O_i's.

Obviously this formulation requires eventual refinement to include consideration of the distribution of benefits across consumers. We are also aware that, for most public bureaus, it is extremely difficult to obtain reliable and valid measures of outputs, of the values per unit of output, and even for the components of a cost function. The difficulty of measuring costs is not, however, avoided by using the concept of fiscal residuum. While measuring the outputs of some public bureaus is extraordinarily difficult given the public good nature of those outputs, we feel it is essential to develop a conceptual performance measure that captures some aspects of the meaning of the term, "the public interest."

Having defined the benefits residuum as an entry in bureaucratic objective functions, we hasten to add that there are likely to be additional entries in those functions as well. Bureau chiefs are assumed capable of

diverting some bureau resources to personal rewards, perhaps additional perquisites of office, increased status among his or her peers, or, like some characterizations of private sector managers, an easy life. The extent to which such capabilities are exercised depends upon bureaucrats' motivations and the constraints they confront. We turn next to some models of bureaucratic motivation and then to constraints.

Models of Bureaucratic Managers

Two polar models of bureaucratic managers can be advanced. One of these, a "selfless bureaucrat" model, seems implicit in much of the literature of public administration. The second polar model suggests a totally rapacious bureaucrat, striving to capture all possible surpluses for himself and for those whom he must satisfy to maintain his position. We suspect that neither polar model captures the reality of bureaucratic motivations, but briefly review each before offering an alternative.

The selfless bureaucrat in our terminology is solely interested in maximizing the benefits residuum. The selfless bureaucat knows citizens' preferences sufficiently well to be able to define "the public interest." He or she pursues that public interest with single-minded devotion. Using this model one can safely dispense with controls aimed at monitoring or constraining bureau chiefs' exercises of discretion. Thus when one hears of reforms designed to strengthen hierarchical control in public bureaus or, even more, when one hears of reforms designed to reduce political control or "interference" in the operation of public bureaus, one can assume the proponents are employing a selfless bureaucrat model. By eliminating unnecessary impediments, they would argue, bureau chiefs are freed to maximize net benefits to their consumers. This selfless bureaucrat model might be viewed as the equivalent of the pure profit maximizing model in the private sector. Given that profit maximization is not the sole strategy adopted in the private sector unless heads of firms are driven to it by the rigor of a competitive market, it seems naive to assume complete selflessness in the public sector unless comparable mechanisms exist to force such single-minded bureaucratic pursuit. We see few institutional arrangements in the public sector that carry such force.

On the other hand, total rapaciousness seems naive also. Models of this nature posit budget maximization as the sole strategy, where bureau chiefs attempt to extract the full consumer surplus available to a consuming community by increasing budgets (and output) to the maximum obtainable. We disagree with such a formulation on two different grounds. First, it seems to us to entail too much work on the part of bureaucrats and bureaucratic employees. Increasing output beyond some point, even if it were to lead to subsequently higher budgets, is likely to increase the workload of bureau managers beyond their preferred levels. Second, we perceive bureau chiefs, at least at the local level, as taking some pride in "doing a good job." Part of that good job, we believe, includes supplying net benefits to their consumers.

Because of our reservations about either polar model of bureaucratic managers, we choose to use a utility maximizing model, with objective function entries drawn from both polar models. Our utility maximizing model is basically a managerial discretion model, modified from the work of Williamson (1964), Pondy (1968), and Orzechowski (1977). We assume that bureau managers derive utility from the benefits residuum and from resources devoted to bureau personnel. The utility derived from a positive benefits residuum comes from the increased probability of job retention and advancement associated with a positive consumer surplus among the constituents of the providing organization and from personal satisfaction

with serving the public well. Many local urban service bureau managers live in the community they serve and consume the output of their own bureau. We feel it is reasonable to assume that a local public service bureau chief will want to gain confidence and appreciation from citizens served and from friends, family, and neighbors for creating a positive consumer surplus. We also feel it is reasonable that a bureau chief would be willing to trade-off some utility derived from the benefits residuum for increased utility derived from investment in bureau personnel.

In terms of total numbers of employees, a bureau chief's own salary and status are usually a positive monotonic function of the size of the bureau. However, the chief may derive even more satisfaction from investments in specialized personnel and staff assigned to help with the administrative load. In regard to hospitals, Lee (1972: 85) has argued that "inputs are used as status symbols, or, in other words, the pattern of input utilization defines the status group to which a hospital belongs." He also argues that hospital managers participate in a "keep up with the Jones's game" in that the "desired inputs of, say the ith hospital is assumed to be a function of the inputs utilized by other hospitals" (Lee, 1972: 85). If this same game characterizes urban police departments, which we think is the case, departments and their managers derive considerable status and recognition for investing in specialized personnel. Having their own homicide investigation bureau, bad check or arson team, dispatch facility, crime lab, and entry-level training academy adds to the status, and thus, the utility of an urban police chief. The sworn personnel assigned to administration significantly lighten the workload of a chief and also contribute to his utility.

Formally, the model we adopt posits that managers of urban police departments derive utility from both the benefits residuum and the number of staff or specialized personnel. In other words, the manager attempts to:

$$\text{Max: } U = (NP,\ BR) \quad \text{Subject to } TC \leq B \qquad (10)$$

The constraint is that costs cannot exceed the approved budget. Equation (10) can be rewritten to:

$$\text{Max: } U = (NP,\ \sum_{1} v_i O_i - w_p P - w_{np} NP - w_c C - r_a A - \sum_{i} r_{o_i} O_i - FC) \qquad (11)$$

Figure 1 shows the relationship between benefits residuum, bureau output, and specialized bureau personnel. For this graphical presentation, all output variety is collapsed to a single output index. In the figure we construct iso-benefits residuum contours for all staff and output combinations. Along the ridge line (L), the marginal effects on the benefit residuum of an increase in output holding staff constant is zero. Along (D) the marginal effect on the benefits residuum of increasing staff while holding output constant is zero. At (K) where the two ridge lines intersect, the optimum staff-output combination exists in terms of the benefits residuum. By slicing through the iso-benefits residuum contour in Figure 1 along the ridge line (L), we obtain the possibility curve relating staff to the benefits residuum shown in Figure 2. Points (K) and (A) are the same in both diagrams. We can use the possibility curve and axes of Figure 2 to illustrate the difference between a selfless bureaucrat's objective function and preferred staff and benefits combination and the objective function and preferred position of our utility maximizing bureaucrat.

The selfless bureaucrat would derive no utility from the employment of staff specialists per se. His or her single-minded preference for benefits residuum is illustrated in Figure 2 by the parallel horizontal lines labelled I_{sb1} and I_{sb2}. The line I_{sb2} represents a higher level of utility for the selfless bureaucrat and is achieved solely through

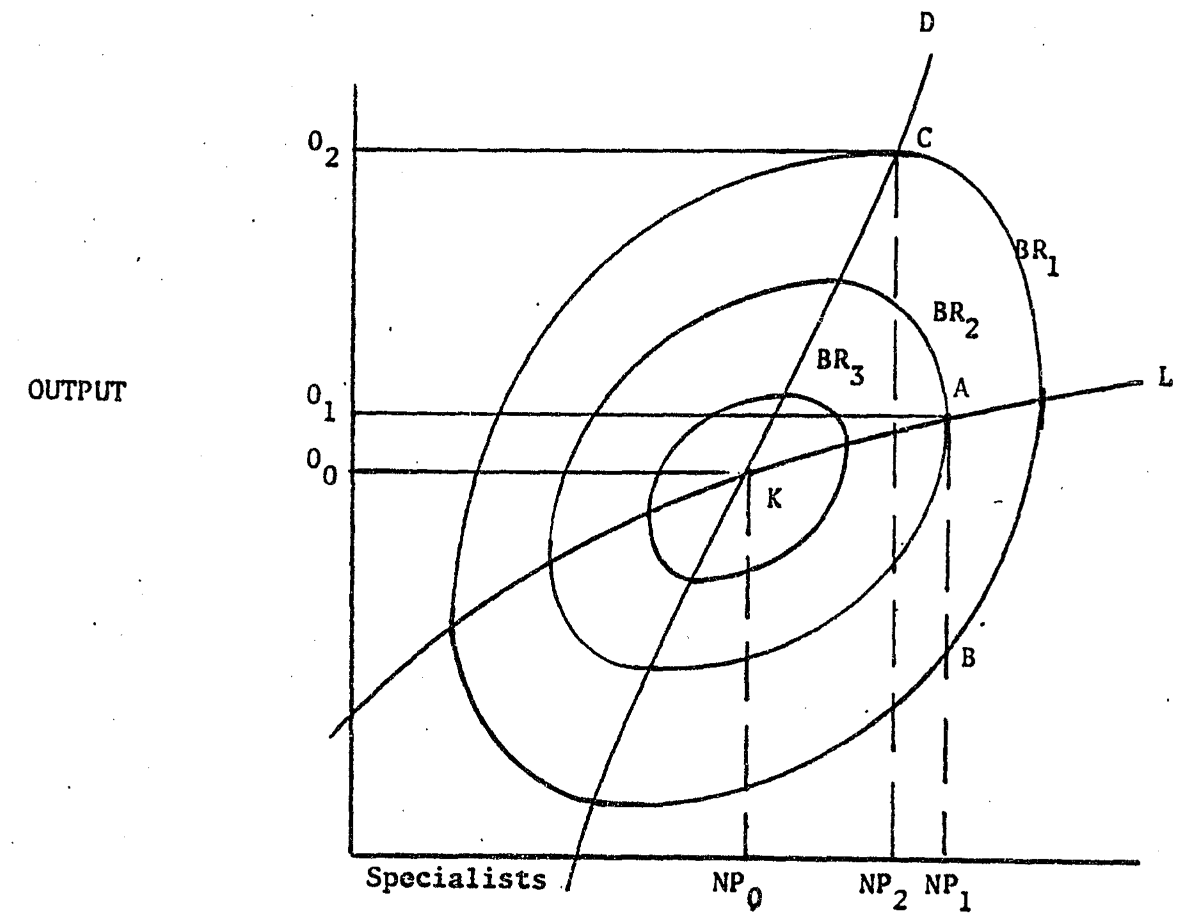


Figure 1

Iso-Benefits Residuum Contours

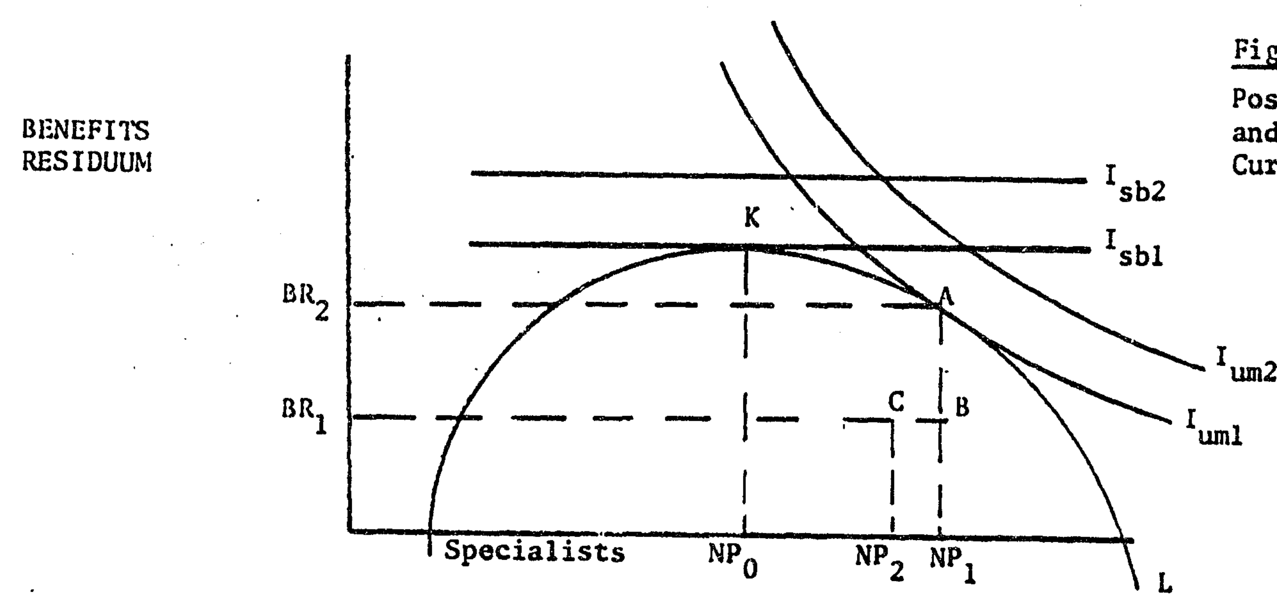


Figure 2

Possibility and Utility Curves

increases in benefits residuum. Thus, a selfless (and efficient) bureau manager would select the output-staff combination which produces (K) on the benefits residuum. If the selfless bureaucrat were not efficient, some point below the frontier possibility curve would be chosen. In this case a reduction in the benefits residuum would result from the bureau chief's inefficiency rather than from the objectives sought by the bureau chief.

While the indifference curves of a "selfless" bureaucrat would be horizontal, the indifference curves of a utility maximizing bureau would tip to the right and be convex to the origin as shown by the two curves labeled I_{um1} and I_{um2}. Thus, the utility maximizing and efficient bureau chief would select a staff-output combination at A, with a lower level of net benefits for the community. If the bureau manager were not technically efficient, more staff than NP_0 would be employed, but less output would be produced. An inefficient but utility maximizing bureau chief might end up at Point (B) in both figures where the number of staff personnel is more than NP_0 and the increase beyond optimal levels (for the community) does not even produce the most output for the combination of input resources. Any utilization of staff larger than NP_0 can be considered an investment in excess staff from the perspective of consumers. The amount of excess staff can also be considered a measure of the extent of managerial discretion that can be exercised.

The extent to which utility maximizing bureaucrats are able to exercise their preferences without constraint is a function of the institutional arrangements in which they operate. Any point along the possibility frontier from (K) to (A), or interior points to that section of the frontier might result from bargaining between bureaucrats and officials of providing organizations.

Bureau managers must negotiate with providers on a regular but infrequent basis (sometimes once a year), for authorization to spend a lump sum over a defined period of time. Niskanen argues that the nature of the relationship between a bureau and the officials of a providing organization frequently approximates that of a bilateral monopoly. Given that officials of providing organizations frequently have no other potential supplier of bureau services, bureau managers may gain the "same type of bargaining power as a profit-seeking monopoly that discriminates among customers or that presents the market with an all-or-nothing choice (Niskanen, 1971: 25)." If officials of the providing organization are unwilling to forego the bureau's services, they may be at a disadvantage in the negotiation over the amount of budget to be approved for a bureau. If the bureau is able to conceal information about its production and cost functions while obtaining substantial information about the demand characteristics of members of the provider's constituency, the bureau chief's capacity to confront providers with a take-it-or-leave-it proposition is enhanced (Stockfisch, 1976). This capacity is further enhanced when no competitive or potentially competitive proposals are forthcoming, either from alternative suppliers or from comparative analyses by providers of the proposals offered and accepted in other, similar situations. Of course institutional arrangements linking the provider and the bureau (and, where applicable, other potential suppliers) will affect the relative bargaining strengths of each. The situation is not fully determinate as in Niskanen's first model (1971), but rather will depend on these relative strengths (Breton and Wintrobe, 1975; Niskanen, 1975). In addition, the role of constituents or consumers as they constrain provider behavior through elections and other means must be considered in fully developed models (see MacKay and Weaver, 1978, and Langbein, 1980, for models incorporating consumers as voters).

Interorganizational Arrangements and Bureau Supply

Just as we think that the linkages between managers of a particular bureau, the officials of the providing organization with which bureau officials negotiate, and the citizen-consumers who are the constituents of the providing organization will have significant influences on the bureau's supply, so we believe that the interorganizational structure of service delivery in a particular area will affect the supply by all bureaus in the area. Interorganizational structure for the delivery of police services varies dramatically from area to area across the United States and also from service to service within the broad police service rubric (Ostrom, Parks, and Whitaker, 1978). These diverse service delivery arrangements for policing in this country afford us fertile grounds for empirically testing effects of differing interorganizational structures. In this section we first describe a bit of the variation in service delivery structures for policing that we have found and then suggest how some structural differences might affect bureau supply.

In 1974 and 1975 we conducted a census of all organizations supplying public police services in 85 metropolitan areas (SMSAs) across the country. We found that the number of suppliers of patrol service in the 85 areas ranged from one single supplier in Meriden, Connecticut, to in excess of 90 suppliers in Paterson-Clifton-Passaic, New Jersey. Fewer than one half of the areas had a single supplier of patrol service for as much as 50 percent of the population. The median number of producers of general area patrol in an SMSA was 13. The existence of multiple producers of this service in the metropolitan areas meant that, at least in theory, providers could obtain comparative data to weigh against the budget and output performance of their current producing bureau and might even replace their current supplier with one or more competitors.

We posit that the number of patrol producers (multiplicity) in an SMSA should affect the level of information that citizens as consumers and officials as providers have concerning the relative benefits residuum produced by a bureau that supplies their jurisdiction with patrol services. In a metropolitan area where there are many different producers (high multiplicity), citizens obtain information about comparative performance in several ways. Simply driving through the metropolitan area provides regular information about patrol density and the extent and style of enforcement in different jurisdictions. If a citizen or a member of his or her immediate family receives a traffic ticket in two different jurisdictions, an opportunity exists to compare directly the fairness, courtesy, and honesty of officers working in different jurisdictions. Most citizens in a metropolitan area with many jurisdictions know residents living in many different jurisdictions. Informal discussions of such personal events as being victimized, calling the police for assistance, or getting a ticket often occur among friends.

Public officials in a metropolitan area with high multiplicity are apt to be better informed about comparative performance levels in the metropolitan area also. Citizens who are unhappy with their own police and who know that their friends and neighbors receive a better level of service are more apt to call their elected officials than citizens living in a low multiplicity area who have no way to compare the service they receive with that of other jurisdictions. Further, if city managers and/or mayors in the metropolitan area meet regularly, they can exchange relevant input and output information that helps each of them in their bargaining with police chiefs. The relative monopoly over information that Niskanen posits is reduced in a metropolitan area with a large number of producers.

Thus, police chiefs operating in metropolitan areas with high levels of multiplicity are more exposed to removal if they increase staff and other input variables beyond the level at which more effective departments in the metropolitan areas operate.

We expect that the lower costs of monitoring police department performance in metropolitan areas with high multiplicity will affect the shape of the indifference curves for bureau managers operating in those areas. In Figure 3 we illustrate how this might operate by examining the relationship between increasing the staff of a producing agency in any particular jurisdiction and the risk of exposure to the monitoring activities of citizens and provider organization officials. As the risk of exposure increases, bureau managers confront such sanctions as severe criticism in the local press or by the city-council, or even being fired for the relatively high costs of policing in one jurisdiction when compared to similar jurisdictions in the immediate vicinity. Where the benefits residuum is increased by adding to staff, there is little or no risk of exposure to sponsor monitoring activities. However, the risk of exposure should rise exponentially once the benefits residuum maximum combination of staff and other inputs has been surpassed. We assume that this "risk" curve is affected by the level of multiplicity in the metropolitan area. In areas characterized by large numbers of other producers, the risk is higher at all levels of staff beyond the optimal number for the benefits residuum (NP_0). This contextual effect of multiplicity on the relationship between size of staff and risk of exposure to monitoring activities should be reflected in the size of staff in police departments serving similar populations in metropolitan areas varying from low to high multiplicity. This results from the posited changes in the shape of managers' indifference curves as shown in Figure 4.

Using our assumption that police bureau chiefs are utility maximizers, Figure 4 shows the indifference curves for bureau chiefs in two different types of metropolitan areas (SMSAs). In low multiplicity SMSAs, the information available for monitoring a chief's performance in trading off benefits residuum and staff is low. Therefore, chiefs in such areas may be more free to indulge a preference function with a shift toward increased staff at the expense of some benefits residuum. In high multiplicity SMSAs we posit that more information for monitoring will be available and, thus, chiefs will be forced to develop preference functions closer to that of the selfless bureaucrat, emphasizing benefits residuum more than their counterparts in areas with less information. One might argue that chiefs learn the risk of exposure and adjust their preference functions accordingly. As an alternative for those who wish to keep any individual chief's preferences constant, one can envision chiefs whose preferences are inappropriate given the level of information available for monitoring being replaced by a new top administrator whose preference function is appropriate to the situation. Police chiefs in the United States typically have quite short tenures (approximately 2 years on average) so this adjustment process should occur with fair rapidity.

The usefulness of utility maximizing assumptions (as opposed to assuming a single maximand such as budget, output, or benefits residuum) depends upon being able to specify the arguments in the utility functions of managers and to identify instances where managers' costs of pursuing different combinations of those entries vary from situation to situation (Alchian, 1965). We have specified two entries in such functions for police managers, the benefits residuum or net community benefits supplied by their efforts and the number of specialized personnel employed by the agency. We have also

Figure 3

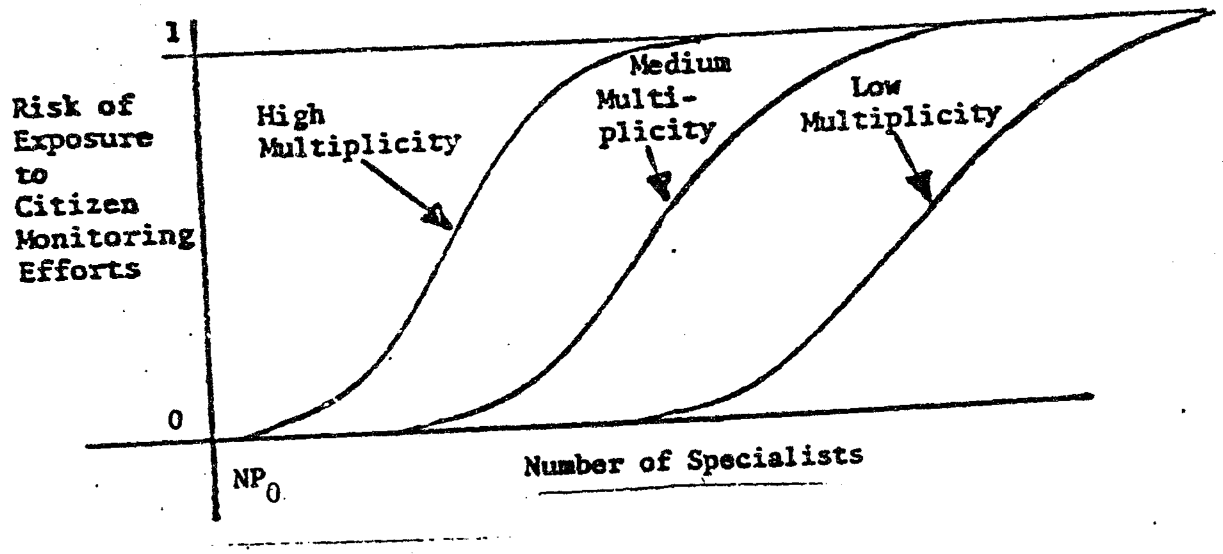
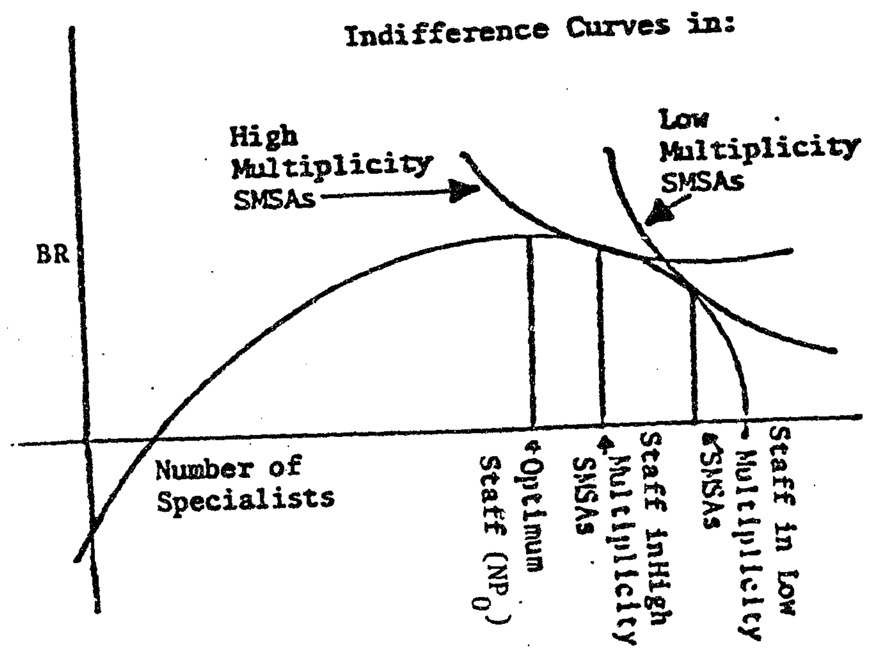


Figure 4



identified variations in interorganizational structure that will, we believe, provide some empirical explanation for differences in the choice of benefits residuum-specialists combinations across police agencies in different structures. We anticipate that police agencies operating in structures that reduce the cost of monitoring by citizens and by officials of provider organizations will choose to employ relatively fewer officers in specialized assignments, thus, moving to the left on their benefits-specialists possibility curves.

Estimating Average Police Service Production Functions

In this section we estimate production function for two common police outputs. After estimating those functions across all departments in our sample, we will examine the functions for departments in differently structured SMSAs. We will see if the modes of production differ as we might expect from our discussion of utility maximizing bureau chiefs and the differences in their exposure to monitoring in different service delivery structures.

The two outputs we have chosen for anlaysis are: (1) police response capacity -- measured here by the average number of police patrol cars on the street,[4] and (2) arrests for serious crimes -- measured by the number of Part I crimes cleared by arrest in a year. While police produce many varied outputs (see Goldstein, 1977), these two are certainly among the most important to consumers and to the police themselves. A high proportion of police resources are typically devoted to maintaining an on-street presence and response capacity. Commonly recommended measures of police performance usually include response time as a measure of capacity and clearances as a measure of the effectiveness of departmental investigative activities. Thus, we feel comfortable in using these output measures for our initial estimations.

Our output specification in Equation (3) has been modified for estimation purposes by taking an internal production strategy choice into account. That choice, whether to deploy officers in one- or in two-officer units, has been the subject of substantial controversy and some research in police circles (Boydstun, Sherry, and Moelter, 1977). It should certainly have an influence on the output elasticities for officers assigned to patrol duties. To take this into account, we restate the output elasticity for officers with patrol assignment, a, in the following way:

$$a = a_0 + a_1 PCT2, \text{ where} \qquad (15)$$

PCT2 = The percent of on-street patrol units with two officers assigned.

Our estimates for these production functions are based on data from a large sample of municipal police departments. The personnel and response capacity data were collected during 1974 and 1975 from police agencies operating in 85 metropolitan areas across the United States. These areas, while excluding the very largest, were otherwise representative of most American metropolitan areas (Ostrom, Parks, and Whitaker, 1978 -- five additional metropolitan areas located in Wisconsin have been added to the sample reported there). The clearance data were obtained from the Federal Bureau of Investigation's Uniform Crime Reporting Section, and are the clearances reported by the municipal departments in our sample for the year 1973. The departments used to develop the estimates in the following tables ranged in size from 5 sworn officers to more than 1,300 sworn. The departments, thus, sample from the spectrum of full-time departments in the United States with the exception of 15 to 20 very large agencies.

In our statement of Equation (3) we explicitly accounted for the joint production of police outputs by including terms for the quantities of all

other outputs produced into the equation for the quantity of any single output. Where this simultaneity of output equations exists, ordinary least squares regression estimates of the functional coefficients may be biased. Unfortunately, we do not have simultaneous equation estimates for those coefficients available at this time. In order to report some empirical findings, our discussion in this section relies on estimates derived from single equation least squares techniques.

The estimated coefficients for the response capacity production function are shown in Table 1. We have used the percent of the population served by each police agency whose income was below the poverty level in 1970 as an indicator for service conditions. The total number of crimes cleared by arrest is included in the equation to account for the possibility of joint production. The production function coefficients have been estimated in four size ranges to examine whether production relationships vary with police agency scale.

The estimates show that elasticities for officers assigned to patrol duties are generally much higher than those for officers with nonpatrol assignments or for civilian employees. This is quite reasonable given the output in question. However, both nonpatrol assignments and the use of civilians do make stronger contributions among the largest departments. The estimates for the clearance by arrest coefficients show virtually no effects on response capacity in the smaller departments, and a positive effect in the largest. This may indicate economies of joint production in the larger agencies.

We can exercise these production function estimates by computing the output expected from typical departments in each size range. Table 2 presents some data for doing this and the results using these typical figures.

Table 1

Cobb-Douglas Production Function Coefficients for Response Capacity Output -- Average Number of Cars on the Street

Factors	5 to 30 Officers	31 to 75 Officers	76 to 150 Officers	More than 150 Officers
	Scale of Production -- Police Departments That Employ:			
Patrol Officers				
a_0	.519[#] (.047)[@]	.660 (.154)	.407 (.201)	.444 (.173)
a_1 (% in 2-Officer Cars)	-.0008 (.0002)	-.0018 (.0003)	-.0013 (.0004)	-.0001 (.0002)
Nonpatrol Officers				
b	.021 (.005)	.099 (.083)	-.004 (.130)	.240 (.155)
Civilians				
c	.013 (.004)	.026 (.017)	-.043 (.078)	.236 (.135)
Automobiles				
d	.187 (.042)	.079 (.080)	.479 (.139)	-.324 (.158)
Clearances by Arrest				
g	-.005 (.011)	-.008 (.037)	.053 (.082)	.285 (.109)
Percent of Population That is Below Poverty Level				
h	-.036 (.016)	.048 (.024)	.012 (.031)	.080 (.108)
Scale Factor				
ln k	-.549 (.105)	-.956 (.455)	-1.192 (.981)	-2.158 (.563)
R^2	.72	.59	.46	.81
(N)	(271)	(72)	(57)	(40)

[#]Unstandardized regression coefficient [@]Standard error of coefficient

The estimates appear quite reasonable and suggest decreasing returns to scale for response capacity.

Table 2

Predicted Response Capacity Output for Typical Police Agencies -- Average Number of Cars on the Street

Factors	Total Number of Sworn Officers: (20)	(50)	(100)	(300)
Patrol Officers*	15	34	63	168
% in Two-Officer Cars	0	16	22	40
Nonpatrol Officers	5	16	37	132
Civilians	3	9	20	75
Automobiles	4	10	25	75
Clearances by Arrest	40	100	200	600
Percent of Population that is Below Poverty Level	15	15	15	15
Predicted Average Response Capacity (Cars)	2.8	6.5	8.1	18.7

*Sworn officer assignments for typical departments from Ostrom, Parks, and Whitaker, 1978, 89-93.

The scale efficiency factor, computed as the sum of the output elasticities minus 1.0 -- see Equation (8) above -- is -.26 for the smaller departments. It is -.17 for the next largest, -.21 for departments with 76 to 150 officers, and -.41 for the largest departments. These findings indicate that, for purposes of supplying response capacity, departments could economize by reducing their level of output with correspondingly larger reductions in the required inputs. This is particularly true for the largest departments.

These decreasing returns with scale of response capacity output are in part a reflection of additional types of outputs as department size increases. If the number of output types and the scale of a particular output (such as response capacity) are correlated, then a part of the decreasing returns found in a single output estimation reflects this correlation. We hope to have multioutput, simultaneous equation estimates in the near future to further examine the scale efficiency question.

The estimated coefficient a_1 in these equations shows the expected effect on the output elasticity for patrol officers of placing them in two-officer cars. The effect appears quite large for medium-sized police departments, but relatively small for the largest. This probably reflects the fact that a much higher proportion of large departments currently use two-officer cars (Ostrom, Parks, and Whitaker, 1978: 93) and, thus, may have developed compensating strategies to reduce the marginal productivity impact.

Table 3 presents the coefficient estimates for our clearance production function. For the small- and medium-sized departments, officers assigned to patrol duties and automobiles tend to dominate the output function. In the larger departments, however, other inputs become equally or more important, including the effect of higher response capacities. The scale efficiency parameter for these output functions is quite different from that for response. It is equal to .08 for the smaller departments, .80 for the medium-sized, -.36 for departments of 76 to 150 officers, and -.50 for the largest departments. This indicates that it could be economical for the production of clearances to increase the scale of that production among small- and medium-sized departments, perhaps by combining the investigative activities of two or more agencies. The predicted numbers of clearances

Table 3

Cobb-Douglas Production Function Coefficients for Crime Fighting Output -- Total Clearances by Arrest

Factors	Scale of Production -- Police Departments That Employ: 5 to 30 Officers	31 to 75 Officers	76 to 150 Officers	More than 150 Officers
Patrol Officers				
a_0	.439[#] (.306)[@]	1.38 (.57)	-.087 (.362)	-.269 (.276)
a_1 (% in 2-Officer Cars)	-.0014 (.0012)	-.0028 (.0014)	.0008 (.0007)	-.0003 (.0004)
Nonpatrol Officers				
b	.051 (.028)	.097 (.285)	-.244 (.222)	.115 (.236)
Civilians				
c	.027 (.024)	.015 (.058)	.455 (.119)	.074 (.207)
Automobiles				
d	.568 (.234)	.356 (.271)	.502 (.258)	.597 (.224)
Average Number of Cars on the Street				
g	-.153 (.332)	-.097 (.426)	.158 (.246)	.620 (.236)
Percent of Population That is Below Poverty Level				
h	.242 (.086)	.136 (.083)	-.038 (.054)	.038 (.160)
Scale Factor				
ln k	1.324 (.588)	-.580 (1.601)	4.383 (1.606)	3.361 (.808)
R^2	.17	.33	.40	.79
(N)	(271)	(72)	(57)	(40)

[#]Unstandardized regression coefficient [@]Standard error of coefficient for typical departments show this same pattern. Using the same inputs as in Table 2, the estimated clearances per sworn officer are 2.5, 3.2, 6.1, and 4.9 across the increasingly larger departments.

Effects of Service Delivery Structure with Average Production Functions

We expected to find that bureau chiefs in SMSAs characterized by higher levels of multiplicity would choose more efficient input mixes than would their colleagues in lower multiplicity areas. This expectation resulted from our understanding of the risks of exposure of, for example, overstaffing with specialized officers. We argued that these risks would be larger in high multiplicity areas where information on production opportunities might be more available. In our discussion of bureau production functions, we presented the result that efficient producers would use input factors in proportion to their output elasticities and inversely proportional to their wage ratios -- see Equation (7) at page 9. Given the output elasticities computed in the preceding section and the knowledge that the wages of nonpatrol officers almost universally exceed those of patrol officers, we can put some bounds on the efficient mix of patrol and nonpatrol officers for departments in each size range. Assuming for the moment that wages are the same for both types of officers, the optimal proportion of officers with patrol assignments is related to the elasticities as:

$$p = R/(1+R), \text{ where} \qquad (16)$$

$$R = (a_0 - a_1 \times PCT2)/b, \qquad (17)$$

and a_0, a_1, and b are coefficients for the estimated output functions. If the wage rates are not equal, but rather nonpatrol officers receive higher wages, the p is a lower bound on the efficient proportion of officers given patrol assignments.

To see whether metropolitan service delivery structures affect officer assignment choices in the way expected, we computed the average proportion of officers with patrol assignments in departments located in differently organized SMSAs. By comparing those proportions to the proportion appropriate to the size of the department, we can perhaps say something about the efficiency of the mixes chosen. Table 4 presents data for this exploration.

Table 4

Efficient Sworn Officer Assignments and Actual Assignments in Different Service Delivery Structures

	Scale of Production -- Police Departments That Employ:			
	5 to 30 Officers	31 to 75 Officers	76 to 150 Officers	More than 150 Officers
P_{min}-minimum Proportion of Officers with Patrol Assignments	.90	.86	Could Not Be Computed	.65
P_{actual}-proportion of Officers with Patrol Assignments				
SMSA Multiplicity (Quartiles)				
1 to 7	.77	.69	.62	.54
8 to 13	.78	.73	.66	.60
14 to 20	.81	.66	.61	.56
21 or more	.78	.69	.60	.56

There is very little, if any, patterning of assignments with respect to the multiplicity of patrol producers in an SMSA. From these data we cannot say that departments in high multiplicity SMSAs were any more likely to choose efficient input mixes than were those in low multiplicity SMSAs. This negative finding also held true when we examined differences in relative multiplicity, the number of patrol producers per 100,000 SMSA population, and differences in dominance, the proportion of the SMSA population served by the patrol producer with the largest serviced population. Rather than concluding that service delivery structures have no effect, however, we explored an alternative way of estimating efficient service production functions and found some interesting effects.

Alternative Estimating Techniques

Use of the Cobb-Douglas specification for modeling the production function for response capacity may be appropriate in that many estimated production functions have been found to resemble this specification. But using OLS estimation techniques with the logged equation and data from a large sample of departments raise other problems. It produces estimates for the coefficients of the model based upon data from departments in the data set that may or may not have chosen technically efficient production strategies (efficient in the sense of maximizing output for a given set of inputs). If all departments confronted equal incentives toward production efficiency this would not be a major problem. However, as one of our major research interests is the factors that contribute to choice of efficient production strategies, the use of average production functions is undesirable on this score.

M. J. Farrell (1957) offers a method for determining efficient production functions from observations on the inputs and outputs of many different firms. In concept the method is simple, One plots all combinations of inputs leading to a given level of output and then determines the innermost envelope of those points (the set of points closest to the origin on each of the input dimensions) for each of the isoquants of interest.

Charnes, Cooper, and Rhodes (1978) offer a related method and show how it can be used with readily available linear programming algorithms.

Figure 5 shows a graphical representation of the method using only two inputs and one output. This restriction enables the presentation of two dimensional envelopes. Three such envelopes are shown in the Figure, one for departments deploying from 1 to 5 patrol cars, a second for departments deploying 6 to 15 cars, and a third for departments deploying 16 to 30 cars. These envelopes were determined by dividing the number of sworn officers assigned to patrol and the number assigned to other duties by the number of patrol cars on the street at 10 pm, and then plotting these ratios against one another for clusters of departments with given ranges of outputs.[5] The plots were then scanned and the innermost envelopes drawn for each cluster. Within each cluster, those departments that lie to the right of the envelope employ either more sworn officers in patrol for each car on the street, more sworn officers with assignments other than patrol for each car, or a combination of more officers with both patrol and nonpatrol assignments for each car on the street.

Decreasing returns to scale are shown for response capacity output using the frontier function approach, just as they were found using average functions. That is, the frontier production functions represent the minimum combinations of inputs that were observed to produce given outputs. The frontiers for higher levels of output lie to the right of the frontiers for lower levels of output, thus, showing the need for more than proportional increases in inputs as the scale of this particular output increases. Of course, as with the average functions discussed above, a portion of the decreasing returns to scale in the production of response capacity shown here must be attributed to the increased likelihood that departments are producing other types of outputs as the scale of their response capacity increases. We will investigate this phenomena using the frontier function technique (generalized to n inputs and m outputs -- see Farrell, 1957; Charnes, Cooper, and Rhodes, 1978) as our work in this area progresses.

Two average production function estimates are shown in Figure 5 for comparison purposes.[6] The average estimate for 3 cars shows that approximately 4 sworn officers with patrol assignment are required for each car on the street, a figure quite close to that developed in various engineering estimates of officer requirements (e.g., Kapsch, 1970; Misner, 1960). The frontier function for 3 cars on the street indicates the possibility of achieving this output with, for example, 2.5 sworn officers per car on the street -- 1.5 with patrol assignment and 1 with a nonpatrol assignment. Thus, average estimates indicate a requirement of at least 12 officers to put 3 cars on the street, while frontier estimates indicate this could be done with as few as 7 to 8 officers. The average estimates for deploying 10 cars indicate a minimum of about 80 officers required, while the frontier estimates show the possibility of deploying 10 cars with approximately 40 officers. These are indeed wide differences in input requirements.

Figure 6 presents frontier production possibility estimates to show how metropolitan structure may, in fact, make a difference. These frontiers show the maximum combinations of clearances by arrest and cars on patrol (both standardized by the number of sworn officers) that were obtained by departments in metropolitan areas with differing amounts of multiplicity.[7] These frontiers show the trade-off possibilities for response capacity and clearances among the most efficient departments, with the normal concave shape that one expects. They also show a significant upward shift in output possibilities as the number of patrol producers in a metropolitan

area increases. The most efficient producers supply more output for given inputs in high multiplicity SMSAs than do the efficient producers in lower multiplicity areas.

We believe that the frontier curves of Figures 5 and 6 are leading us to some interesting insights. It is likely that the shifts shown in Figure 6 result in part from the presence of more departments of smaller sizes in SMSAs with higher multiplicity. Diseconomies among the fewer (and larger) departments in low multiplicity areas could explain their curve locations. At the same time it seems reasonable that a part of the shifts are attributable to the differences in availability of information in differently structured areas. We are currently working to split apart these two reinforcing tendencies, attempting to better specify the effects of service delivery structures.

Where Do We Go From Here?

In this paper we have proposed several models that we believe are useful in exploring police and other public agency performance. These models include an output or production function model, a model of bureau supply that suggests utility maximizing bureau chiefs may trade-off increments of net community benefits for status-raising choices of personnel deployment, and a model of some ways that interorganizational structure might influence such trade-offs. Our continuing efforts employing these models are several. At a methodological level we are developing alternative specifications for our output functions and moving toward simultaneous estimation of functions for multiple outputs. This will, we hope, give

give us a much better view of the scale efficiency question for agencies of differing size. In those alternative specifications we are introducing indicators for the number and different types of services supplied by the agencies so as to avoid biases introduced by their omission. We are also conducting some exploratory work comparing those departments falling close to the frontier output function estimates to those that lie closer to the average functions or to their right. This will help us to identify additional factors for improving our output model specifications, as well as providing some early evidence with respect to metropolitan structural effects.

At the theoretical level we are moving toward more complete specifications of intra- and interorganizational influences on bureau decision making. At the intraorganizational level we will be attempting to specify likely influences from unionization, where we think this will lead to a preference for particular types of inputs and assignments, and influences from differing rank structures characterizing the relative length of chains of command. At the interorganizational level we are attempting to specify influences from additional structural factors, particularly the extent to which service supply in an area is dominated by a single bureau, and influences from the diversity in service conditions among and within jurisdictions in each area.

We have presented our models and empirical results to date as a work in progress. We hope that our efforts to consider interorganizational influences will encourage other scholars to include consideration of such influences in their own theoretical and empirical work. We welcome suggestions and critiques on our work to this point as well as for further efforts.

Footnotes

[1]The use of producer, consumer, and provider comes from Ostrom, Parks, and Whitaker (1974, 1978) and, earlier, from Ostrom, Tiebout, and Warren (1961). Savas (1978) uses the terms provider, consumer, and arranger where we use producer, consumer, and provider.

[2]In recent years many more sophisticated production function specifications have been developed (Intriligator, 1978, Chapter 8). Most empirical estimation has, however, employed the Cobb-Douglas specification. Most readers will find such a specification more familiar. We intend to explore alternative specifications as our work continues, with particular attention to the general form of translog functions because of their capacity to fit virtually any form (Christensen, Jörgenson, and Lau, 1973).

[3]We use our own term, provider, here, rather than the term, sponsor, which is more common in the literature presenting formal models of bureaucratic behavior.

[4]Response capacity is actually a more complex function of the number of cars available and the volume and distribution of service requests and the time to service each. We are currently developing queueing theory based estimates of response capacity, including the number of response units and call volume estimates derived from functions representing the service conditions in each police agency's response area.

[5]The method of standardization, dividing inputs by outputs requires the assumption of constant returns to scale. We make such an assumption within somewhat arbitrary classifications of output level but, by examining several such classifications, we are able to consider variations in scale returns.

[6]These average function curves are computed from estimates made prior to the final data runs for this paper. Thus, they are not consistent with the data in Table 1. Average curves based on the function estimates in that Table would lie somewhat to the right of those shown here.

[7]In reality these are "nearly maximum" combinations. In examining the plots for these relationships, some points appeared aberrant in that they were far removed from any similar points. Such aberrant points were excluded in drawing the frontiers.

Bibliography

Alchian, Armen A. (1950) "Uncertainty, Evolution and Economic Theory." Journal of Political Economy, Vol. 58 (June), 211-221.

____________ (1965) "The Basis of Some Recent Advances in the Theory of Management of the Firm." Journal of Industrial Economics, (November), 30-41.

Bain, Joe S. (1959) Industrial Organization. New York, New York: John Wiley & Sons, Inc.

__________, Richard E. Caves, and Julius Margolis (1966) Northern California's Water Industry. Baltimore, Maryland: The Johns Hopkins Press.

Boydstun, John E., Michael E. Sherry, and Nicholas P. Moelter (1977) Patrol Staffing in San Diego: One- or Two-Officer Units. Washington, D.C.: Police Foundation.

Breton, Albert and Ronald Wintrobe (1975) "The Equilibrium Size of a Budget-Maximizing Bureau: A Note on Niskanen's Theory of Bureaucracy." Journal of Political Economy, Vol. 83, No. 1 (February), 195-207.

Charnes, A., W. W. Cooper, and E. Rhodes (1978) "Measuring the Efficiency of Decision Making Units." European Journal of Operations Research, Vol. 2 (September), 429-444.

Christensen, L.R., W. Jörgenson, and L.J. Lau (1973) "Transcendental Logarithmic Production Frontiers." Review of Economics and Statistics, Vol. 55 (February), 28-45.

Coase, R. H. (1937) "The Nature of the Firm." Economica, Vol. 4 (November), 386-405.

DeAlessi, Louis (1969) "Implications of Property Rights for Government Investment Choices." American Economic Review, (March).

Downs, Anthony (1967) Inside Bureaucracy. Boston, Massachusetts: Little, Brown, & Co.

Farrell, M. J. (1957) "The Measurement of Productive Efficiency." Journal of the Royal Statistical Society, Series A, 120, Part 3, 253-281.

Goldstein, Herman (1977) Policing a Free Society. Cambridge, Massachusetts: Ballinger Publishing Company.

Intriligator, Michael D. (1978) "Applications to Firms; Production Functions and Cost Functions." In Michael D. Intriligator, ed. Econometric Models, Techniques, and Applications. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 251-

Kapsch, Stefan (1970) Minnesota Police Organization and Community Resource Allocation. St. Paul, Minnesota: Minnesota State Planning Agency.

Langbein, Laura I. (1980) "Production or Perquisites in Public Bureaus." Prepared for the 1980 meeting of the Public Choice Society, San Francisco, California, March.

Latsis, Spiro J. (1972) "Situational Determinism in Economics." British Journal of Philosophical Science, Vol. 23, 207-245.

Lee, Maw Lin (1972) "Interdependent Behavior and Resource Misallocation in Hospital Care Production." Review of Social Economy, (March), 84-96.

Leibenstein, Harvey (1976) Beyond Economic Man. Cambridge, Massachusetts: Harvard University Press.

Levin, Henry M. (1976) "Concepts of Economic Efficiency and Educational Production." In J. T. Froomkin, D. T. Jamison and Roy Radner, eds. Education as an Industry. Cambridge, Massachusetts: Ballinger Publishing Company, 149-198.

Lewellen, W. G. and B. Huntsman (1970) "Managerial Pay and Corporate Performance." American Economic Review, Vol. (September), 710-720.

MacKay, R. J. and C. L. Weaver (1978) "Monopoly Bureaus and Fiscal Outcomes: Deductive Models and Implications for Reform." In G. Tullock and R. E. Wagner, eds. Policy Analysis and Deductive Reasoning. Lexington, Massachusetts: Lexington Books.

Margolis, Julius (1964) "The Structure of Government and Public Investment." American Economic Review, Vol. 54 (May), 236-242.

Marris, Robin (1964) The Economic Theory of "Managerial" Capitalism. New York, New York: Basic Books, Inc.

McKean, Roland N. (1964) "Divergences Between Individual and Total Costs Within Government." American Economic Review, Vol. 54 (May), 243-249.

Mique, Jean-Luc and Gerard Belanger (1974) "Toward a General Theory of Managerial Discretion." Public Choice, Vol. 17 (Spring), 27-47.

Misner, Gordon E. (1960) "Recent Developments in Metropolitan Law Enforcement." Journal of Criminal Law, Criminology, and Police Science, 50-51 (Jan.-Feb.), 497-508 and (July-Aug.), 265-272.

Moran, R.A. (1977) "The Importance of Economic Criteria to Agency Administrators." Evaluation Quarterly, Vol. 1 (February), 173-182.

Niskanen, William A., Jr. (1971) Bureaucracy and Representative Government. Chicago, Illinois: Aldine-Atherton.

__________ (1975) "Bureaucrats and Politicians." Journal of Law and Economics, Vol. 18 ( ), 617-643.

Orzechowski, William (1977) "Economic Models of Bureaucracy: Survey, Extensions, and Evidence." In Thomas E. Borcherding, ed. Budgets and Bureaucrats. The Sources of Government Growth. Durham, North Carolina: Duke University Press.

Ostrom, Vincent and Elinor Ostrom (1965) "A Behavioral Approach to the Study of Intergovernmental Relations." The Annals of the American Academy of Political and Social Science, Vol. 359 (May), 137-146.

__________, Charles M. Tiebout, and Robert Warren (1961) "The Organization of Governments in Metropolitan Areas," American Political Science Review, 55 (December), 831-842.

__________ and Gordon P. Whitaker (1974) "Community Control and Governmental Responsiveness: The Case of Police in Black Communities." In David Rogers and Willis Hawley, eds. Improving the Quality of Urban Management, Vol. 8, Urban Affairs Annual Reviews. Beverly Hills, California: Sage Publications, 303-334.

__________, Roger B. Parks, and Gordon P. Whitaker (1978) Patterns of Metropolitan Policing. Cambridge, Massachusetts: Ballinger Publishing Co.

Parkinson, C. Northcote (1957) Parkinson's Law and Other's Studies in Studies in Administration. New York, New York: Ballantine Books.

Pondy, Louis R. (1968) "Effects of Size, Complexity, and Ownership on Administrative Intensity." Administrative Science Quarterly, Vol. 14, 47-60.

Robinson, E. A. G. (1958) The Structure of Competitive Industry. Chicago, Illinois: The University of Chicago Press.

Savas, E. S. (1978) "The Institutional Structure of Local Government Services: A Conceptual Model." Public Administration Review, Vol. 5 (September/October), 412-419.

Stockfisch, J. A. (1976) Analysis of Bureaucratic Behavior: The Ill-Defined Production Process. RAND P-5591, Santa Monica, California: The RAND Corporation.

Tullock, Gordon (1965) The Politics of Bureaucracy. Washington, D.C.: Public Affairs Press.

Walters, A. A. (1963) "A Review of Production and Cost Functions." A Review of Economics.

END