10. LORRAINE EDEN AND MELVILLE L. MCMILLAN

Local Public Goods: Shoup Revisited

Since the 1954 publication of Samuelson's "The Pure Theory of Public Expenditure," a considerable literature has emerged which focuses on the concept and characteristics of public goods. The purpose of our paper is to review the literature on local public goods (LPGs), a subset of public goods characterized by benefits that taper off spatially. The intent is to situate this literature in relation to Carl Sumner Shoup's work on particular government services, notably fire prevention and police protection services. Our concern is with the issues identified by Shoup, as early as 1964, and their subsequent treatment (or nontreatment) in the literature.

Shoup's writings focus on the characteristics of public services, the definition and measurement of output of individual services, the slopes and shapes of the cost functions as service level, population, and area vary, and the distribution and incidence of the benefits from such services. Much of his work deals with fire protection and crime prevention services. Some of these areas have been actively pursued by subsequent researchers; others have not. Alternatively, some of the current areas of interest in local public finance have not been addressed by Shoup.[1] While his taxonomy of public services has not supplanted the more popular Musgrave-Samuelson approach, Shoup's concern with the particularities of individual government-provided services has generated a large amount of subsequent research.

Our paper looks at developments in those areas of local public finance which we believe link most closely to Shoup's contributions. The first section of the paper focuses on the characteristics of LPGs, and is followed by sections on defining and measuring output, cost functions, and distributional issues. The fifth section summarizes these results and addresses the question: have the issues raised by Shoup been successfully answered, or are there areas that remain to be addressed? We conclude that empirical and theoretical studies of local public goods have just begun to handle the four issues Shoup addressed twenty years earlier and that much work remains to be done.

The Characteristics of Local Public Goods

The Musgrave-Samuelson Approach

Two of the key early writers on the characteristics of public goods were Paul Samuelson (1954, 1955, 1958a) and Richard Musgrave (1959, 1969). The widely quoted definition of a public good as one where "each individual's consumption of such a good leads to no subtraction from other individuals' consumption of that good" (Samuelson 1954:179) defines the first characteristic, *nonrivalness in consumption*. In his 1959 book Richard Musgrave refers to social wants as "those wants satisfied by services that must be consumed in equal amounts by all" (8). He argues that *nonexcludability*, that is, the inability of private providers to exclude potential consumers through the price mechanism (and thus be unable to require consumers to reveal their preferences), implies that social wants must be met through public provision. This thinking is formalized in R. Musgrave (1969), which identifies the two characteristics of social goods as nonrivalness in consumption and nonexcludability from consumption. Nonexcludability means that the market is likely to be an inefficient means of providing such commodities; however, whether public provision is necessary or superior to the market depends on the circumstances (see Head 1972).

Pure public goods have both characteristics. Figure 1(A), based on Stiglitz (1986:103) and Musgrave, Musgrave, and Bird (1987:45), shows a box with combinations of two characteristics, rivalness and excludability.[2] Point a represents the polar case of a pure public good; point c represents the polar case of a private good.

Where local public goods like fire, police, and recreational services fit within the box depends partly on the author and the circumstances. In practice fire and police services, and often recreational services, are publicly produced and provided free of direct charge at the municipal level to residents of the jurisdiction. Stiglitz (1986:100) situates fire protection toward the lower right-hand corner, that is, the marginal cost of protecting an additional household in a community is low, but it is

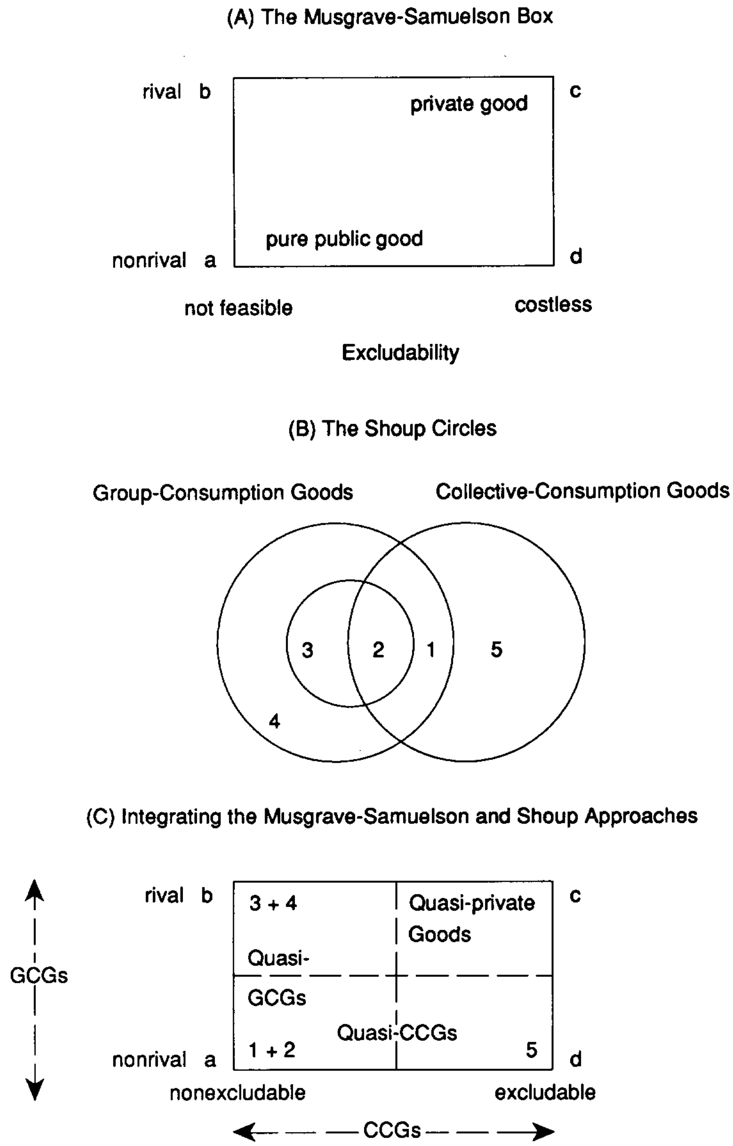

Figure 1: The Musgrave-Samuelson and Shoup Approaches to Public Goods

feasible to exclude the household from the service. Police protection, however, may fall toward the lower left. Although the cost of protecting another household is low, it is not feasible to exclude it (fully) from the benefits which police services afford the community (even if direct services were not provided to the specific unit).

The Shoup Approach to Public Goods

Shoup uses a different taxonomy for public goods than the Musgrave-Samuelson one outlined above.[3] Commodities are distributed free of charge by the government for four general reasons, according to Shoup: preservation of the nation-state against aggression, group-consumption goods, redistribution in kind, and miscellaneous reasons such as difficulty in measuring output and uninsurable costs (Shoup 1969b:65). In general only services are freely distributed because goods can be resold whereas few services are resalable.

Of these four reasons Shoup concentrates on the second: the *group-consumption good* (GCG) (see also Shoup 1974). A GCG is defined as a good or service that can be supplied more efficiently through a nonmarketing method than by a rationing (price or nonprice) method. The nonmarketing or group method means that the GCG is supplied simultaneously to all members of the group, and exclusion is not feasible. The key criterion for choosing between the two methods is *efficiency*, defined as production and distribution at lowest cost including the cost of resources used in operating the method. Efficiency must be defined in incremental terms as the cost of an additional unit of the service in per capita terms. It is possible for a particular good to switch from being a GCG to a non-GCG, depending on the relative costs of the two methods at different levels of supply.

Within the group-consumption good framework Shoup addresses the specific characteristics of individual GCGs. When the service is government provided, it is available within a given area from one or more points of input (e.g., a fire station). If the benefits taper off as the recipient is located further away from the point of input, different households receive differing amounts of the service depending on their location within the jurisdiction. This type of GCG is now identified as a *local public good.* Shoup notes that the point of input can be mobile (e.g., patrol cars) and there may be several input points (e.g., recreation centers). Spillovers of benefits between units can also occur.

While nonexcludability of a GCG means that discrimination between individual households or firms is impossible (or at least less costly than exclusion), discrimination at the level of the group may be possible. Such *intragroup discrimination* may be involuntary or deliberate on the part of the local government. For example, involuntary discrimination between rich and poor neighborhoods can happen even though equal numbers of police are provided to both areas because the inputs (number of police) do not produce equal benefits (the same risk of becoming a victim of crime) in rich and poor areas. Deliberate discrimination between groups can occur if distinct subunits exist within a jurisdiction and the local government discriminates in the level of service provided to different areas (e.g., by deliberately providing higher service levels in wealthy areas).

Separate from the concept of a GCG is a *collective-consumption good* (CCG), which Shoup defines as one that, if supplied to one person, can be supplied to additional people at zero incremental cost. Thus the total cost of supplying a given level of service of a CCG to each household remains unchanged as the size of the consuming group expands. A CCG may be excludable. Shoup illustrates the range of excludability with theaters and mosquito abatement. A CCG thus has the characteristic of nonrivalness of consumption but may be either excludable or nonexcludable, whereas a GCG has the nonexcludability characteristic but may be either rival or nonrival.

The distinction between GCGs and CCGs is illustrated in figure 1(B) which is based on Shoup (1969b:73, and chap. 5). The left-hand circle, composed of segments 1, 2, 3, and 4, represents GCGs (nonexcludable public services); the right-hand circle, composed of areas 1, 2, and 5, is CCGs (nonrival public services). The GCG circle contains a smaller circle, areas 2 plus 3, where intragroup discrimination is feasible; outside this circle, in area 1 plus 4, intragroup discrimination is not possible. Outside both GCG-CCG circles are goods that are excludable and rival, that is, private goods.

The individual segments can be explained as follows. Where the two circles overlap (categories 1 and 2), the services are nonrival and nonexcludable; this is the polar case of a public good. In category 1, pure public goods where intragroup discrimination is not feasible, Shoup includes military expenditures, public health, space research, contract enforcement, and externalities. Shoup speculates that category 2, pure public goods where intragroup discrimination is feasible, is nearly empty. Category 3, GCGs with feasible intragroup discrimination, includes fire and police protection, street maintenance, and flood control. Shoup conjectures that category 4, GCGs with nonfeasible intragroup discrimination, is also empty. He hypothesizes that category 5, excludable CCGs, includes education and medicine, waste removal, and recreational services.[4]

Shoup (1969b, 1974) clearly situates both fire and police protection as publicly provided, rival services with feasible intragroup discrimination (area 3). He argues that the marginal cost of extending fire and police services to an additional consumer is nonzero so the services are rival. Once the service is provided by the municipality, it is not feasible to exclude individual households from consumption. Intragroup discrimination is possible, but not at the level of the individual household or firm. Shoup positions recreational services, however, in category 5 as nonrival goods that are most efficiently produced and distributed by private firms.[5] Walsh in this volume, on the other hand, argues that the case for private as opposed to public provision is not clear and that work remains to be done in this area.

The Musgrave-Samuelson box in figure 1(A) has an advantage over Shoup's circles in that movements within the box can be interpreted as corresponding to more or less rivalness and/or excludability; that is, rivalness increases in a northward movement, excludability in a westward movement. Such directions cannot be read from figure 1(B). Figure 1(A), however, does not allow for the intragroup discrimination/nondiscrimination split that exists in Shoup's circles. Shoup's approach also encourages thinking about the relative mix of GCGs and CCGs through changing the sizes of three circles and their degrees of overlapping. For example, a small right circle implies fewer joint goods and a smaller public sector; a large overlap between the two main circles increases the range of pure public goods. The Musgrave-Samuelson box, on the other hand, encourages linear thinking along the diagonal of the box.

We compare Shoup's GCG-CCG taxonomy with the Musgrave-Samuelson public goods taxonomy in figure 1(C). The comparison between the Shoup circles and the standard box treatment of public goods does not appear to have been made before. The comparison is a bit forced since exclusion in the Shoup taxonomy refers to all nonmarketing modes including queuing and rationing, whereas in the Musgrave-Samuelson approach exclusion normally refers to sale through the private market. Categories 1 plus 2 in figure 1(B) correspond to point a (pure public goods), categories 3 plus 4 (rival, nonexcludable services) to point b, and category 5 (excludable, nonrival services) to point d.[6] Private goods, the

Rec

area outside the circles in figure 1(B), are represented as the single point c in figure 1(C). Thus the polar case of a group-consumption good can be translated as the vertical line ab in figure 1(C); the collective-consumption good, as the horizontal line ad. The Shoup focus on the polar GCG-CCG cases (the ab-ad lines) is therefore quite different from the Musgrave-Samuelson focus on the polar case of a pure public good, the single point a.

In the Musgrave-Samuelson approach everything in the box other than points a and c is a grey area of quasi-public goods of varying degrees of rivalness and excludability. In the Shoup taxonomy, since GCGs are based on the criterion that nonexclusionary supply is less costly than rationing, we can interpret quasi-GCGs as the left half of the box in 1(C). Similarly, CCGs can be extended to include public services with positive, but low marginal costs per user. Quasi-CCGs are thus the bottom half of the box. Since these areas overlap, the lower left-hand quadrant can be seen as quasi-GCG/CCG cases. This leaves the last quadrant as quasi-private goods with c as the polar case.

Shoup's definitions of group-consumption goods and collective-consumption goods have not supplanted the more popular Musgrave-Samuelson rivalness-nonexcludability approach. It is interesting to speculate as to why the Shoup taxonomy has not been more widely used. Perhaps the terminology is somewhat confusing (i.e., the difference between the terms "group" and "collective" is not obvious). In addition Shoup's definition of a group-consumption good depends on both the costs of production and distribution of the service; whereas subsequent literature has separated the issue of production (government versus private sector) from the question of provision (financing). The definitions of nonexcludability also differ; Shoup's includes both price and nonprice forms of rationing (e.g., he defines public education as excludable because it is rationed). Most authors (see Head 1972) use excludability to mean exclusion via the price mechanism only. In the following section we review the recent literature on the characteristics of local public goods that deals with the issues Shoup raised.

Recent Developments: The Characteristics of LPGs

Shoup's definition of group-consumption goods depends on the relative efficiency in the production and distribution of local public goods under the marketing versus nonmarketing (i.e., nonrationed) modes. In order to determine which method is most efficient, costs of private production, public production, private sale, public distribution free of charge, public distribution with user fees, and public distribution with rationing must be computed and compared. The most efficient method is the least cost combination of production and distribution methods for a given level of service. Given the data requirements, it is not surprising that little work has been done on nonexcludability as defined by Shoup.[7] Most work in local public finance has focused on rivalness, that is, on estimating points in a vertical direction in the Musgrave-Samuelson box. The methodology of Bergstrom and Goodman (BG, 1973), Borcherding and Deacon (BD, 1972), and other papers using their approach is the basis for much of our review.

The Basic Rivalness Model and Results

It was only with the pathbreaking BG-BD papers that estimates of the rivalness of public services became available. Both papers introduce into their models of the demand for public output a term to allow for the potential rivalness or congestion of the publicly produced good. They assume that the amount of the public output to the individual (q) is determined by the amount of output produced (Q) and the number (N) sharing that output as follows:

$$q = Q/N^{\alpha} \tag{1}$$

If $\alpha = 0$, $q = Q$ and the public output is a pure public good. If $\alpha = 1$, $q = Q/N$ and the public output has private good characteristics in that each person benefits from only a per capita share of the total output. Intermediate values imply quasi publicness/privateness. The rivalness factor α is referred to variously as the crowding, capturability, congestion, and/or publicness parameter. It is the congestion elasticity and measures the percentage change in public output Q needed to maintain constant at q the amount of the public service benefiting each beneficiary for a given percentage change in N.

The congestion parameter is then introduced into a utility maximizing demand model. Continuing to follow the Bergstrom and Goodman specification, an individual's utility (u_i) is a function of private goods consumed (X_i) and q; that is, $u_i\,(X_i, q)$. The person's budget constraint is:

$$X_i + r_i\, p\, Q \leq Y_i \tag{2}$$

where the price of X_i is one, r_i is the tax share of the person, and p is the unit cost of Q. Recognizing that $Q = N^{\alpha} q$, the budget constraint is:

$$X_i + r_i \, p \, N^{\alpha} q \leq Y_i \qquad (2.1)$$

where $r_i \, p \, N^{\alpha}$ is the price per q-unit to the individual. Assuming demand is Cobb-Douglas with constant price and income elasticities, the demand for q is:

$$q = (r_i \, p \, N^{\alpha})^{\delta} \, Y_i^{\sigma} \qquad (3)$$

and the demand for Q is:

$$Q = (r_i \, p)^{\delta} \, N^{\alpha\,(\delta+1)} \, Y_i^{\sigma} \qquad (4)$$

In application the form of the estimated equation is:

$$\ln E = \phi \ln N + \delta \ln r + \delta \ln p + \sigma \ln Y + \Sigma \, \beta_i X_i \qquad (5)$$

where E is municipal expenditures (pQ) on a particular function, ϕ equals α $(\delta + 1)$, and X_i denote other socioeconomic characteristics influencing expenditures and conditioning utility. This approach assumes that public decisions reflect the median voter's preferences so that r and Y represent the median voter's tax share and income. The value of the publicness parameter α can be derived from the estimated values of ϕ and δ. The results from the BG-BD demand studies yield an α value close to unity.

Since the BD-BG papers, many investigators have incorporated the publicness specification into their own models. The prevailing empirical result is that α is close to one, indicating that publicly provided services are subject to congestion. The rivalness parameter values derived for fire and police services specifically tend to be somewhat more diverse (see below).

Specification of the Congestion Term

The BD-BG specification of the congestion term is a simple one, $q = Q/N^{\alpha}$. An implication of this specification is that for $\alpha > 0$ congestion is decreasing at the margin. That is, $dq/dN < 0$ and $d^2q/dN^2 > 0$. Furthermore, the congestion elasticity, $\alpha = (dq/dN)/(q/N)$, is a constant. Such a form may be unduly restrictive. Indeed, one can readily think of public services such as roads and swimming pools for which congestion could be increasing at the margin as population or the number of users expands.

Concern for the restrictiveness of the publicness function N^{α} has led various investigators to consider alternative forms. Edwards (1986) considers several models allowing for variable congestion elasticities and for increasing as well as decreasing marginal congestion. Where permitted, decreasing marginal congestion performed better than other specifications. Hayes and Slottje (1987) test two alternate congestion relationships—($N^{\alpha + \phi \ln N}$), which reduces to the original N^{α} when $\phi = 0$, and $e^{\beta N}$, which allows congestion elasticities to vary with population—against the simple N^{α} specification. The former is found superior to the exponential $e^{\beta N}$ form, but the BG-BD specification dominates both as statistical tests cannot reject the hypothesis that $\phi = 0$. These results suggest that the original BD-BG congestion parameter is quite robust.

Instead of the usual single-equation approach to public expenditures, Hayes (1985, 1986) and Hayes and Slottje (1987) use a simultaneous system of equations to study the demand for fire protection, police services, and other local public goods. Hayes (1986) allows α to be influenced by demographic variables—metropolitan (versus nonmetropolitan) and growing (versus nongrowing) cities. These demographic variables do not significantly influence the congestion effects.

Output and the Publicness Parameter

Several researchers have suggested that the observed nonpublicness of some government services may be partly a measurement anomaly. Most of the forementioned studies circumvent (albeit neatly) the problem of actually measuring the output of the public service itself. As Bradford, Malt, and Oates (1969) and Shoup have noted, estimates of rivalness are likely to be inaccurate when the underlying output of LPGs is incorrectly measured. In this section we look at the impacts of output definition, size and number of facilities, and varying distribution of benefit levels on congestion measures.

Brueckner (1981) examines congestion in local fire protection services by estimating the reduction in expected fire losses resulting from expenditures on fire suppression capability. The expected reduction in losses is based on fire insurance premiums which vary among communities depending upon the Insurance Services Office rating of the quality of fire protection in each. He finds congestion elasticities that range from

$-.1143$ to $-.2379$. These estimates indicate a high degree of publicness for fire protection.

Craig (1987) models police protection as a process by which police labor inputs produce a clearance rate and clearances deter crime and produce safety. The rivalness specification Craig selects is $(\mu - N)^{\sigma}$ which permits increasing rather than decreasing marginal congestion as population increases. If $N = \mu$, congestion stalls production and nothing is produced; for smaller N, output is realized and increases as N falls. The congestion elasticity, $[-\sigma N/(\mu - N)]$, varies with population, and, for $0 < \sigma < 1$, increasing marginal congestion costs occur and congestion elasticity becomes, absolutely, larger. Using data for Baltimore police beats, Craig finds that the total elasticity of the final output (safety) to increases in N is small (-0.016) when N is low; however, as population increases, the congestion elasticity reaches -1.0, implying privateness. Craig also estimates his model with the conventional BD-BG specification. The estimated elasticity (-0.047) implies a high degree of publicness, but the implications of the two models for larger beats differ somewhat. No test of the alternative specifications is provided.

McGreer (1989) compares the performance of the BD-BG with Craig's and Edwards's exponential specification of the congestion term using Australian local government data for recreation and culture, roads, and total municipal expenditures. A test of alternative specifications reveals that the BD-BG specification dominated the Craig version in each case. The BD-BG model also appears superior to the exponential form for the first and last categories while neither dominate for road maintenance.

Other economists have been concerned that the data studied and the approach of the analysis tended to mask the underlying publicness of LPGs. Many services, for example, fire and police stations and cultural and recreational facilities, have limited service areas and/or populations. In larger communities a single facility may be inadequate to meet local demands. If within a single community multiple facilities are required, replication of units will make numbers and cost correspond closely to population (unless there are substantial economies in overhead) even though the service at the individual facility level has publicness characteristics. To test this idea McMillan, Wilson, and Arthur (1981) estimate congestion parameters, using the BG model, for small and large municipalities (less and more than 10,000 population). They find that the congestion elasticities for general services, recreation and culture, and fire protection in communities under 10,000 persons are typically about one-half as large as those for the over 10,000 group which (with the exception of that for recreation and culture) approaches one.

Communities with larger populations may be able to provide a greater range of services than smaller jurisdictions. This is what Oates (1988a) refers to as the "zoo effect," since only larger municipalities can usually afford, for example, zoos. If expenditures for services expand with population, because the range of services expands, estimates of the congestion parameter not accounting for this change will be upward biased. This situation may partly explain the observed nonpublicness of local public goods. Some evidence of Oates's contended zoo effect is provided by McMillan (1989) in a study of fire protection in Ontario municipalities. The quality of fire protection service (Oates's range of service) is measured by the fire protection grade assigned by the Fire Underwriters' Survey. The grade of service tends to improve with population size. McMillan finds that inclusion of fire protection grade in a standard demand equation reduces the congestion parameter substantially.

If the personal level of service can vary among residents, the distribution of q can affect the publicness parameter. Gramlich and Rubinfeld (1982) introduce this possibility into their demand model by assuming that the level of q residents realize is conditioned by a relative income term. The distribution of q favors the rich or the poor as the income distribution parameter is greater or less than zero. Since values for this parameter estimated from household data exceed zero, this indicates a (perceived) pro-rich distribution of local schooling. The associated estimates of the publicness parameter from jurisdictional and individual household data yield values of about one.[8]

Summary of Evidence on the Congestion Parameter

Columns 5 and 8 in table 1 report the publicness parameter values for police and fire expenditures, respectively, for fifteen different studies published during the 1972–87 period. The statistically significant estimates range from 1.02 to 1.63 for police services and from 0.39 to 4.27 for fire services. Shoup's placement of fire and police services as GCGs is neither confirmed nor denied by this evidence; however, the estimates do suggest that the marginal cost of provision is nonzero.

Table 1: Estimates of Price, Income, and Congestion Elasticities[a]

Study	Price Term	Police Protection			Fire Protection		
		price	income	α	price	income	α
Bergstrom/Goodman (1973)	tax	(0.25)	0.71	1.07	NA	NA	NA
Borcherding/Deacon (1972)	wage	(0.97)	0.81	1.02	—	0.88	1.01
Brueckner (1981)	wage	NA	NA	NA	0.03	—	(0.24)[f]
Deacon (1978)	index	(0.76)	—	—	(0.99)	—	—
Ehrenberg (1973) low	wage	(0.01)	NA	NA	(0.23)	NA	NA
high		(0.35)			(0.31)		
Gonzalez/Mehay (1985)	none	NA	0.23	NA	NA	0.69	NA
Gramlich/Rubinfeld (1982)	wage	(0.06)[b]	NA	NA	(0.06)[b]	—	NA
Hayes (1985)	wage	(0.71)	—	1.50	—	—	0.95
Hayes/Slottje (1987)	wage	(1.23)	0.40	—	(1.03)	—	4.27
McMillan/Wilson/Arthur	tax	NA	NA	NA	(0.62)[c]	—	0.39
(1981)							to 0.69[c]
Pack/Pack (1978)	tax	(0.19)	0.53	1.63	(0.12)[d]	0.47[d]	1.32[d]
					(0.21)[e]	(0.82)[e]	1.58[e]
Perkins (1977)	wage	(0.73)	0.21	NA	(0.75)	0.14	NA
Pommerehne/Frey (1976)	tax	NA	NA	NA	(0.33)	NA	0.56
Santerre (1985)	wage	(0.45)	0.49	1.35	(0.48)	—	1.66
Vehorn (1979)	tax	NA	NA	NA	(0.15)	—	1.0
					to (0.88)		to 2.0
Range of values[g] low		(0.01)	0.21	1.02	(0.06)	(0.82)	0.39
high		(1.23)	0.81	1.63	(1.03)	0.88	4.27

a. NA indicates that no elasticity was estimated; a dash (—) that the elasticity did not differ significantly from zero at the 10 percent level or better.
b. Estimated for fire plus police; the t-value is 1.8.
c. For municipalities with less than 10,000 population.
d. For towns of 1,000 to 5,000 population.
e. For towns of 5,000 to 50,000 population.
f. Elasticity of congestion of fire protection.
g. When significantly different from zero.

Defining and Measuring the Output of Local Public Goods

The Shoup Approach to Output Measurement

Shoup (1969b:78) establishes a clear distinction between units of input and units of output of a GCG. He argues that physical units should be measured in three dimensions: time, number of consumers served, and level or intensity of the service. A correct measure of output is in household-weeks of units of service. A *creative service*, a GCG agreeable to household consumption and/or which creates intermediate or final products for firms, can occasionally be measured in this manner (e.g., area size of a park).

Measuring output of a GCG is difficult at the best of times; when the service is a preventive one, the difficulties are worse. A *preventive service* is "one that is not in itself agreeable to consume or receive, but is valued nonetheless because it prevents, or is deemed to prevent, something more disagreeable from happening" (1969b:78). Shoup argues that most GCGs are preventive services, for example, crime reduction, limiting fire damage, and restricting the spread of disease. By its very nature quantity of a preventive service cannot be measured by the amount of input; the only way to measure the level of a preventive service is indirectly by measuring what it fails to prevent. The analysis must be in decrements of service from an unknown total. Shoup suggests that inputs can be distinguished from outputs as follows: if the service is nonmarketable because discrimination at the level of the household is not feasible, anything that is marketable is an input, not an output.

Fire protection services consist basically of limiting damage once a fire has started together with education in and inspection for fire prevention. Shoup argues that the output of fire protection, as a preventive service, must be measured indirectly by the mean and variance of the expected damage from fire over a stated period in a stated area. Output can be measured incrementally by changing the level of fire protection inputs and observing the subsequent fall in fire losses and/or insurance premiums. The physical output of police protection, another preventive service, is best measured by the number of undesirable (weighted) events that are prevented by police inputs (see Shoup 1964, 1969b, 1974, 1988a). Since the end product is the reduction of exposure of the household to crime, output should be measured as the reduction in the probability of loss of a certain dollar amount of property from crime over a certain period due to a rise in police inputs.

Recent Developments: Output Measurement

Shoup clearly identifies a major issue affecting all theoretical and empirical work on local public goods, that is, how to define a physical unit of a publicly provided service. In the absence of a quantifiable measure of output, demand and cost functions for the LPG, benefit distribution and incidence measures, and elasticities of demand and supply cannot be accurately calculated. Much of the early work on local public goods suffers from this deficiency. (See the review in Beaton 1983.) There are, however, a few papers which distinguish output from expenditure.

Shoup's approach is echoed in Bradford, Malt, and Oates's (1969) distinction between C- and D-output, which uses a two-stage production function. In the first stage inputs of labor, capital, and materials are used to produce a vector of directly produced output (D-output), in the form $D = f(I)$, where I is inputs and D is direct output. In the second stage D-output, in conjunction with environmental variables, E, determines the output that affects the consumers' utility functions, C-output. That is, $C = g(D, E) = g[f(I), E]$. For example, in the case of police services, I could include police officers and cars, D the number of blocks under surveillance, E the proportions of residential to business and urban to rural properties, and C the probability that a resident would not be subject to crime.

A few recent papers have been based on the Shoup–Bradford, Malt, and Oates (BMO) approach. Schwab and Zampelli (1987) use this approach to argue that zoning regulations and fiscal migration between communities can affect the production of LPGs through changing population characteristics. They note that the tax price variable common to most median voter analysis is the price of D-output, not C-output, so that all empirical estimates of demand determinants are biased. To test their hypothesis they include income in the production function to act as a proxy for community characteristics. They find that the usual income elasticities underestimate the true values once income is allowed to affect both the demand and supply of LPGs.

Five recent papers have been based on the Shoup–BMO approach to preventive services, four on police and one on fire expenditures. Scicluna, Foot, and Bird (1982) break the BMO distinction between D- and C-output into an intermediate category, D′-output. They argue that C-output can differ from D-output for two reasons, environmental and productivity/efficiency factors. The intermediate category, D′-output, is defined as operational or D-output adjusted for quality differences, but excluding environmental factors; D′-output is thus a measure of output proper. C-output is then D′-output adjusted for environmental factors. The paper uses these distinctions to measure police productivity as distinct from environmental influences.

Craig (1987), cited above, models police protection as a two-equation process by which police labor inputs produce a clearance rate (D-output) and clearances produce safety (C-output) as reflected in a lower crime rate. Both production processes are subject to congestion. Gyimah-Brempong (1987) estimates a multiproduct translog cost function where arrests for each type of crime are the primary output measures, inputs consist of two types of labor (police and civilian labor) and a measure of capital, and cost shares are taken as the value of property stolen relative to total cost. The third paper by Levitt and Joyce (1987) regards arrests as an intermediate output, where final output is crime deterred. The sole fire paper, Brueckner (1981) cited earlier, estimates the reduction in expected fire losses in a community resulting from expenditures on fire suppression capability.

Cost Functions for Local Public Goods

The Shoup Approach to LPG Cost Functions

Table 1 in Shoup (1969b:143) summarizes his hypothesized effects of level of service, population, and area size on the cost functions of producing and distributing group-consumption goods. Category 1 and 2 services (pure public goods) face the standard rising total cost and U-shaped marginal and average cost curves as the level of service increases, holding population and area constant. Increases in population, for a given level of output and area, leave output unchanged, while marginal cost falls because the marginal cost of provision to an additional consumer is zero for a pure public good. Increases in area size, holding population and level of output fixed, vary by function, but generally leave the elasticities unaffected.

However, as population increases, rivalness may increase as public services become congested. The cost of maintaining service quality may increase and/or service quality may diminish; for example, response time increases as new homes are located further from the fire station, or costs rise as increases in density require more sophisticated and more expensive manpower and equipment. Shoup argues that police and streets face rising total, average, and marginal costs over all three variables (level, population, area). Fire protection faces rising costs as the level of service increases, but U-shaped costs for increases in population and area size. In general in category 3, an increase in population, with a fixed amount of input and fixed area, causes a decline in the level of service. Migration between jurisdictions can therefore cause fiscal externalities by affecting per capita cost. An increase in area size, however, with population and service level fixed, causes costs to increase.

The links between population, rivalness and growth of government share are explored in Shoup (1976, 1984). He concludes that little work has been done in this area. Eden (1984) extends Shoup's analysis to examine the impacts of publicness on government share measured in current and constant dollar terms and in an open economy. See Bird (chap. 13, this volume) for additional extensions linking tax reform and government growth.

Recent Development: LPG Cost Functions

Cost of public service studies would often have benefited from Shoup's insights. Many of the analyses examine scale effects in the sense of per capita cost rather than as expenditure per unit of output (E/Q). It is the latter which is of interest, but that approach is complicated by problems of public output measurement. Constant returns to scale is often assumed in recent studies because it is necessary in the BD-BG framework in order to identify the structural parameters (Inman 1979:295–96). Without constant returns only the product, s α, of the returns to scale parameter (s) and the publicness parameter (α) is available from the estimates. Evidence of returns to scale in local public services is mixed.

Fox (1980) reviews studies of size economies in fire and police protection plus other services. Output measures used in those studies, if not simply population, usually include population with some effort to control for service quality. With respect to fire protection Fox concludes that there are some economies of scale for communities up to 10,000 persons but that they are very limited beyond that size. For police protection he finds that unit cost savings seem to exist with size but that the cost savings are offset by the addition of more services. Rider (1979), however, in a study of fire protection in New York that regards scale as a multidimensional factor involving population, area, and workload, obtains a U-shaped average cost curve. Brueckner (1981) finds some evidence of increasing returns to scale in his fire data for one hundred cities.

The results of police protection studies are mixed. Walzer (1972) concludes that when an index of police service is used, a U-shaped cost function is observed, but when population is used to reflect scale, no economies are revealed. Scicluna, Foot, and Bird (1982:269), in their study of police productivity using the concept of D′-output, find evidence of substantial amalgamating of police forces into regional groups. Gyimah-Brempong (1987), using a multiproduct translog cost function where arrests are the primary output measures, rejects the Cobb-Douglas production function specification. For the average police department decreasing returns are observed; the scale of decreasing returns rises as population increases. Also, he finds no evidence of cost advantages to providing a variety of services; that is, there appear to be no economies of scope in public production. Levitt and Joyce (1987) and Carr-Hill and Stern (1973) also observe diseconomies of scale in public services. Eden and Millar (1990) find that smaller municipalities in north Ontario have, *ceteris paribus*, higher exogenous costs (that is, higher fiscal needs) and lower fiscal capacities than their larger, southern Ontario cousins, providing some evidence of size economies.

We conclude that, while progress is being made in the study of economies of scale in local public services, there is as yet no clear evidence as to whether or not constant returns prevail. One consequence is that, without this evidence, caution is advised when interpreting estimates of publicness.

Distribution and Incidence of the Benefits from LPGs

The Shoup Approach to Benefit Incidence and Distribution

Shoup's work on the benefits from LPGs focuses on three areas: measurement of the benefits, intragroup discrimination, and the shifting and incidence of LPGs. Group-consumption goods provide valuable benefits to their consumers. Shoup argues that their value can be measured by the sum of the amount by which households and firms reduce purchases of private competing services, plus savings in insurance, plus the increment households and firms are willing to pay in higher building and land prices to enjoy the GCG (Shoup 1969b:116).[9] If the service is a preventive one that is not prized for its own sake (e.g., fire protection), Shoup suggests that its benefit can be measured by the saving of market expenses and the reduced disutility from fear and inconvenience.

The benefits from GCGs are not necessarily uniform to all consumers within a jurisdiction. Shoup (1969b) argues that equality of inputs across the jurisdiction is no guarantee of *horizontal equity*, equal treatment of equals, because benefits from LPGs depend not only on inputs, but also on the environment in which these inputs are used. Horizontal inequity in the distribution of benefits is usually due to different costs necessary to attain a certain level of service across various subgroups (e.g., urban/rural, residential/business, rich/poor). Such intragroup discrimination may be involuntary or unrecognized since it is due to socioeconomic variables that affect the technology of public service provision.

Shoup (1988a) hypothesizes that economists have ignored the horizontal equity issue for several reasons: preoccupation with the pure public good concept; the difficulty of measuring output of GCGs; preoccupation with vertical, as opposed to horizontal, equity; lack of availability of micro data; lack of a connection between public services and money valuation; little demand for such studies by recipients; preference of social scientists for measuring public inputs, compared to outputs; and stress on equity among input suppliers rather than output consumers. He calls for more work on the horizontal equity impacts of local public goods, arguing that political scientists and lawyers have contributed more to this issue than economists and that economists can still provide a valuable perspective.

Governments distributing services free of charge to households and firms may deliberately discriminate in the level of service provided to subgroups within the jurisdiction. If such intragroup discrimination is feasible, the jurisdiction must consider how to optimally discriminate among the subunits. Different goals will lead to different methods of intragroup differentiation. For example, the government may choose from several goals for fire protection: maximum reduction in the number of fires, maximum protection from becoming a victim of fire, equal protection per dollar's worth of property. Similarly, there are several goals for crime prevention: equal protection per person, maximum prevention of crime, equal work load per police officer (see Shoup, 1964, 1969b, 1974, 1988a).

A third issue related to the benefits from GCGs is the shifting and incidence of such benefits. A free government good exerts a positive income effect on its recipient; this is the impact incidence of the GCG (Shoup 1969b:86). If there are locational differences in benefit levels, market forces will cause the final incidence to differ from its original impact (e.g., higher rents for households living near city amenities). Shoup uses the term *relinquishment* rather than shifting, to stress the loss of GCG benefit as opposed to the removal of a tax burden (1969b:88). Such relinquishment is most likely to occur when government services are locational; that is, enjoyment of the GCG depends on geographic location of the household or firm relative to the service input. The shifting of benefits is manifested by a rise in the price of marketable goods bought by the GCG consumer and/or a fall in the recipient's factor rewards. Some examples are given in Shoup (1969b:92–93; 1988a:9). Shoup (1988a) notes that the concept of relinquishing benefits is unfamiliar to most public finance scholars and is not discussed in most textbooks, whereas expenditure incidence has been extensively analyzed.

Recent Developments: Benefit Shifting and Incidence

Measuring the Benefits from Local Public Goods

The conventional approach to measuring the benefits from the local public sector is to focus on the vertical distribution of income (Dodge 1975; Gillespie 1976; Dahlby 1985). Using reasonable assumptions, these studies divide public outputs into those which are specific to identifiable groups and those which are indivisible. Allocation of the benefits of indivisible services is done according to income, property, per capita, and other bases. Concerns about this methodology stimulated Aaron and McGuire (1970) to propose an alternative, distinguishing between the benefits of public output imputed from the marginal value of the public good and the impact of redistributive transfers resulting from its finance. Though not without fault, this approach has been employed in many studies. For example, Dean (1980) integrates the Aaron and McGuire method with the BD-BG approach to estimate the benefit incidence of Canadian municipal services. See also studies by Martinez-Vazquez (1982), Weicher (1971), Behrman and Craig (1987), and Hewitt (1987).[10]

Another way to estimate public output benefits is to identify households' willingness to pay from differential property values attributable to variations in local public services. Although a common approach for estimating the benefits and costs of local public goods and bads, this technique has rarely been used to study the distributional implications of the local public sector. We are aware of one application: Chaudry-Shah (forthcoming) uses this approach to measure the capitalization effects of public service and residential property tax variations across twenty-seven subcommunities in Edmonton. He then determines annual residential property tax burdens and the annual value of local public sector benefits and relates these to community income levels. He finds that the amount of the annual tax burden increases with income, whereas the annual value of local public service benefits first diminishes and then increases with income.

Intragroup Discrimination

Twenty-five years ago, Shoup (1964) addressed the problem of how a free government service should be distributed. He refers to services such as fire protection and roadways and deals explicitly with police protection. The question is how the supplier with a specific budget should allocate service across service units that differ in terms of incomes, property values, race, age, etc. In the context of his police protection example Shoup discusses three possible criteria: equal crime rates in each area, equal marginal cost of crime prevention, and tangency of the production and social welfare trade-offs among service areas, with the latter as the preferred alternative. He subsequently notes equal work load per patrolman as the criterion actually utilized in New York City.

Until recently, economists have devoted relatively little attention to intragroup discrimination.[11] However, there are a few papers applied to fire and police services. Rider (1979) examines the distribution of fire companies in New York City. He considers as potential objectives: equal travel times to fires, minimum average city-wide travel time, and equal work load plus fire hazard (as proxied by average assessed value per building). Within residential areas Rider finds that work load per unit area and average assessed value are good predictors. Kennett (1982) looks at distribution of police benefits in New York City. He predicts the number of crimes which would result from applying each of three operational allocation criteria: (i) output equalitarian (equal crime rates per precinct), (ii) input equalitarian (equal input per capita), and (iii) maximum output/minimum offenses (equal marginal crime per patrolman). The actual distribution of police resources does not appear to conform to that implied by any of these standards.

Understanding of the distribution of police services has advanced recently as a result of the work of Behrman and Craig (1987) and Craig (1987) whose models incorporate the interdependence of the public good production and allocation decisions. This interdependence is important because the social value of the product the inputs generate depends upon where they are assigned and whom they benefit. As alternative objectives, Behrman and Craig consider service maximization and equal service as two extreme criteria. They also are able to assess whether the supplier's degree of concern for public outputs varies with neighborhood characteristics; for example, whether poor neighborhoods are favored over the rich. They find evidence of a significant degree of inequality aversion. Police are not allocated to maximize safety (crime reduction), but neither are they allocated to achieve equal services in each area—that is, there is an equity-efficiency tradeoff. In addition the provider's concern appears to vary with neighborhood characteristics.

Price and Income Elasticities of Local Public Goods

Although this paper is motivated by other issues, it would be an incomplete commentary on demand for local public good studies without some mention of estimated price and income elasticities. Appendix I gives a brief discussion of the elasticities of police and fire protection, a summary of which is provided in columns 3, 4, 6, and 7 of table 1.

Conclusions

Carl Shoup's writings on local public goods are based on the taxonomy of group-consumption and collective-consumption goods, a framework quite different from the better-known Musgrave-Samuelson approach. While this methodology has not been more widely used, Shoup's work on public goods that are spatially limited is well known. He emphasizes four aspects of LPGs: the definition and measurement of output, their rivalness and nonexcludability characteristics, the cost functions associated with service level, population, and service area, and the distribution and incidence of benefits.

In this paper we have outlined Shoup's views on each of these issues and then surveyed the subsequent literature. In each of the four areas his work has been at the lead in identifying issues and contributing to the understanding of these questions and, as such, has been at the forefront of the debate. In the case of output definition and measurement it is only recently that empirical studies have begun to take account of his insights. The literature on rivalness, while enormous, still suffers from a reliance on total and per capita expenditures to estimate publicness parameters. As Shoup warned in 1969, such reliance can lead to faulty estimates (1969b). Moreover, Shoup's recognition in his concept of GCGs that nonmarket provision and rivalness are not inconsistent avoids some of the confusion and puzzlement found in some interpretations of congestion parameters implying rivalness. Shoup's recognition that rival goods can be legitimately provided by government (as part of GCG) is itself a contribution. Little work has been done on Shoup's definition of nonexcludability, partly because of the difficulty of making the necessary cost comparisons. In terms of cost functions there is a long history of cost function estimation, but the issues Shoup identified are still being wrestled with today. Lastly, the horizontal distribution and incidence of the benefits of LPGs have been little studied, but are becoming increasingly recognized. Also, the study of LPGs has gone in directions Shoup did not pursue (cf. note 1).

We conclude that Shoup offers economists a broad menu of research topics in local public goods, some of which have hardly been sampled and other areas which, despite active investigation, still have major problems to solve. For those of us working in the area of LPGs, this news is both heartening and disheartening: disheartening because it shows how little we have progressed in particular areas; heartening because there is a good deal of scope for independent contributions to an exciting area of research.

Appendix I: Price and Income Elasticities for Local Public Goods

This appendix provides a brief review of the literature on the price and income elasticities of local public goods. Calculation of price elasticities for public outputs is complicated by the appropriate definition of the price term. With the exception of Deacon (1978) who calculates a price index based on all inputs, the studies noted here are all macro models which use either a median voter tax share or a public sector wage level as the price variable (see column 2, table 1).[12]

The price elasticities reported in table 1 are all negative, and though typically inelastic, the range is broad. For police services the estimated price elasticities range from −0.01 to −1.23, with only one value exceeding unity. For fire protection the range of the elasticities is only slightly narrower and the distribution is also rather uniform. The estimated income elasticities for police services are less than one and fit into a somewhat narrower band, about 0.2 to 0.8. For fire protection most of the estimated income elasticities are not significantly different from zero.

To this point there has been no consideration of developments in the literature suggesting that elasticity estimates from the standard median voter model using cross-sectional data may be biased. However, the literature has recently focused on Tiebout bias, a potential bias that can occur because people select their jurisdiction of residence according to their public good preferences in the spirit of the Tiebout (1956) model. This problem was first identified by Goldstein and Pauly (1981). Their model predicts that demand estimated on observed data with fiscal migration overestimates the true income elasticity when account is not taken of other factors affecting public good preferences. Tiebout bias is a potentially serious problem if households' locations are influenced by local public output. Gramlich and Rubinfeld (1982) provide strong evidence of such sorting. Recent papers by Holtz-Eakin (1986), Rubinfeld, Shapiro, and Roberts (1987), and Bergstrom, Roberts, Rubinfeld, and Shapiro (1988) allow for Tiebout bias in their models. While lower elasticities sometimes result, neither income nor price elasticities are uniformly reduced.

While Tiebout bias implies that the conventionally estimated elasticities are too large, Schwab and Zampelli (1987) argue that, because income may affect both supply and demand, the usual income elasticities underestimate true values. The reasoning for this result is that income is related to household or community characteristics that favorably affect the price of certain services; schooling and crime protection are good examples. In this situation the response of expenditure to income depends not only on the income elasticity of demand, but also on the response to the price change due to the income change. Schwab and Zampelli find that the estimated income elasticity of demand increases from 0.59 to 1.20 when income is allowed to be a determinant of price.

Wildasin (forthcoming) also argues that a second bias, in addition to Tiebout bias, exists when public goods are financed with distortionary taxation. He finds that distortionary taxes raise the effective tax price of local public goods to the median voter. The distortion biases the price elasticity toward unity and the income elasticity away from unity when log-linear demand functions for public goods are estimated, with the error in the price elasticity being much larger than in the income elasticity.

Given these potential sources of bias, estimates of price and income elasticities from public sector demand models should be treated carefully.

Notes

We would like to thank Malcolm Gillis, Paul Hobson, Wade Locke, Carl Shoup, and Enid Slack for comments.

1. For example, there is little mention in Shoup's work on local public goods of the optimal size of the government jurisdiction, the distinction between production and provision, tax exporting, or interjurisdictional migration in response to fiscal differentials, all topics of current research.

2. Carl Shoup has pointed out to us that neither Samuelson nor Musgrave explicitly used the box diagram to illustrate the characteristics of social goods, although Musgrave's table clearly leads to the diagram. Both of us have used the box for years in class lectures and would appreciate knowing who was the originator of this technique.

3. Our discussion is based on Shoup (1969b, Chaps. 4, 5, and 21; 1964; 1974; 1988a).

4. There may be some disagreement as to the public services Shoup includes in category 5, since some of them do not appear to meet the zero marginal cost criterion. The Walsh paper in this volume is an analysis of nonrival but excludable services also. Walsh includes services such as cable television, the output of composers, toll bridges, and entertainment. Since our paper emphasizes GCGs, the Walsh paper can be considered a companion piece.

5. The rivalness/nonrivalness of LPGs is a subject of ongoing controversy. Empirical estimates of publicness can provide insight into the interaction between the rivalness and excludability characteristics. Sometimes, however, the empirical estimates conflict with intuitive positioning (see section entitled "Recent Developments" below).

6. We would argue that category 2 services are less than pure public goods due to the discrimination possibility which, in a sense, is a form of exclusion; that is, those services in category 1 are at point a and those in 2 are a bit to the right of point a.

7. Some work has been done on the contracting out of public services to private producers. The tendency toward privatization of LPGs is a recent phenomenon partly encouraged by rising costs of urban public services.

8. Blecha (1987), however, criticizes the Gramlich and Rubinfeld conclusions about publicness. Once the distributive share term is introduced, she notes that α must vary with population size. Thus the effects of the publicness and distributive share parameters cannot be separated. She argues that this bias is large enough to nullify their results.

9. We note that there could be double-counting here as land prices might reflect lower insurance or private service costs; that is, these have been capitalized into the property.

10. One should be cautious about imputing individual benefits from information of income elasticities derived from interjurisdictional data. Results in Gramlich and Rubinfeld (1982) suggest that elasticities of residents within a community may be quite low relative to the interjurisdictional estimate. They also find that the wealthy within a community perceive themselves as getting greater benefits from public outputs.

11. For reflections from other disciplines, see Benson and Lund (1969), Bloch (1974), Lineberry (1977, 1978), and Rich (1982).

12. For discussions of the price term, see Bahl, Johnson, and Wasylenko (1980), Inman (1979), Rubinfeld (1987), and Bergstrom, Rubinfeld, and Shapiro (1982).