Non-maximizing output behavior for firms with a cost-constrained technology

Jos L. T. Blank

Abstract In many public service industries, firms are constrained by a cost (budget) and characterized by non-maximizing output behavior, due to bureaucratic behavior, for instance. This paper proposes a model based on the assumption that firms with a cost constraint do not maximize service levels due to resource preferences. It derives the exact relationships between services delivered, (shadow) input prices, cost constraints, and optimal input quantities. From these relationships, allocative efficiencies, technical efficiencies, output ray elasticities, and marginal cost can easily be derived.

Keywords Indirect output distance function ·
Shadow cost · Shadow prices · Input preference ·
Allocative efficiency

JEL Classification C31

1 Introduction

Cost function estimation is one of the most popular approaches to analyzing a firm's productivity and efficiency, particularly in public service provision. There are innumerable examples of cost function applications (see e.g., the number of references in Blank 2000). A cost function is a dual representation of a production technology. Under the assumption of cost-minimizing behavior at given service levels and resource prices, it provides a mathematical relationship between actual costs on the one

hand, and exogenous service levels and resource prices on the other. However, it is questionable whether the cost function approach provides a sufficiently realistic representation of actual economic behavior and the economic context in which it occurs. In many public service industries, firms are constrained by a (cost) budget rather than by service levels. There is also sizeable body of literature suggesting that public service firms have motives other than minimizing costs or maximizing profits.

Cost-constrained technologies can be represented by a cost indirect output distance (CIOD) function. This cost indirect output distance function represents the largest factor by which service levels can be improved without the cost involved exceeding a given budget. It is a multiple output version of the well-known production function. The CIOD has only been applied in limited empirical research (see e.g., Grosskopf et al. 1997; Hayes et al. 1998; Blank and Merkies 2004). Since the CIOD requires the same type of data as a cost function, it is not clear why this approach has not been applied more widely.

Since the late 1980s, a number of authors have proposed non-cost minimizing behavior, input preference and bureaucratic models (Eakin and Kniesner 1988; Dor et al. 1997; Eakin 1993; Atkinson and Cornwell 1994; Kumbhakar 1997; Maietta 2000; Atkinson and Primont 2002). They all suggest using a shadow cost function, expressed in terms of exogenous service levels and shadow resource prices. Shadow resource prices (internal to the firm) may differ from actual prices. It is assumed that firms minimize (shadow) cost on the basis of these shadow prices. Shadow cost and shadow resource prices are expressed in terms of actual resource prices and actual cost shares and can be estimated. In fact, the goal of the model is the set of (shadow) resource prices that best fits the actual cost and cost shares under the minimizing-cost assumption. Based

J. L. T. Blank (✉)
Institute for Public Sector Efficiency Studies, Delft University of Technology, P.O. Box 5015, Delft 2600, The Netherlands
e-mail: j.l.t.blank@tudelft.nl

on the principle of shadow pricing, Rodrigues-Alvarez and Lovell (2004) derive a system consisting of an input distance function and corresponding resource demand equations. They estimate the structural under- or over-utilization (expense preference) by estimating an extra intercept in each of the resource demand equations.

Although the CIOD satisfies the cost constraint rather than service-level constraints, it still fails to take the input preference behavior into account. The shadow cost function (SCF), on the other hand, deals with the input preference behavior but fails when it comes to cost constraint. The approach proposed by Rodrigues-Alvarez and Lovell (2004) has another limitation. Their approach breaks the direct relationship between the input distance function and the resource demand equations. The result of this is that any relationship between the under- and over-utilization of resource quantities and firm size, relative input prices and fixed capital is ignored. Potential structural misallocations are assumed to be constant across firms and independent of the characteristics of a firm.

In this paper, I suggest combining the cost-constrained technology approach and the shadow-cost approach by deriving a Shadow Cost Indirect Output Distance Function (SCIOD). This paper will derive an exact relationship between services delivered, actual cost, actual cost shares and actual resource prices. This approach also provides separate measures for technical and allocative inefficiencies in a straightforward manner. In the limited number of studies relating to the CIOD approach, this decomposition has been neglected.

In Sect. 2 I will discuss some of the theoretical issues relating to the CIOD and the SCIOD. Mathematical expressions for the SCIOD and the corresponding resource-demand equations are derived. The expressions for the ray output elasticity and marginal cost are also presented. In Sect. 3, a complete system for the SCIOD based on a translog function is derived, consisting of the indirect output distance function and corresponding resource-demand equations. Section 4 includes some comments on the estimation procedure. The conclusions follow in Sect. 5.

2 Theoretical issues

Non-optimal behavior can be modeled by assuming that the firm is virtually observing a set of input (or output) prices that differs from actual prices. Observed input or output quantities are optimal in relation to these virtually observed prices. These virtual prices are referred to as *shadow prices*. The aim is to establish these shadow prices, theoretically and empirically.

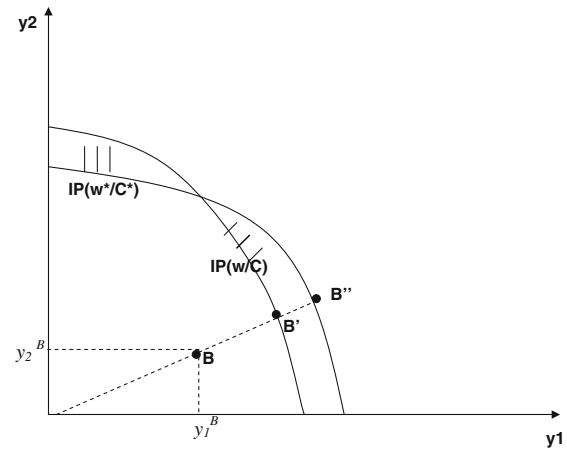Before deriving the SCIOD mathematically, the approach is illustrated by Fig. 1. Figure 1 represents two

Fig. 1 The cost indirect output set and the shadow cost indirect output set

indirect output sets. An indirect output set $IP(w/C)$ is defined as the outer envelope of the direct output sets $P(x)$ for which the cost of the inputs does not exceed a fixed amount of cost C, i.e., $wx \leq C$ (Färe and Primont 1995, p. 83). As relative input prices change, the indirect output set will also change. This alternative vector of input prices is denoted by $w*$, the corresponding cost by $C*$ and the resulting indirect output set by $IP(w*/C*)$.

Firm B is observed and produces (y_1^B, y_2^B). According to the actual input prices and cost, the production of both products of firm B could be expanded by the factor (OB″/ OB). The boundary of the indirect production set functions as the reference set for each observation. At shadow input prices $w*$ (and corresponding cost $C*$) the reference is determined by the shadow cost indirect production set, in which case the expansion factor is reduced to (OB′/OB). The latter ratio is interpreted as technical inefficiency. The ratio (OB″/OB′), which is the result of a reallocation of inputs due to a change in relative prices (i.e., the shadow input prices), can be interpreted as allocative inefficiency. Shadow input prices are established by searching for the vector of (shadow) input prices which minimizes the expansion factor (OB′/OB).

A cost-constrained output technology is defined as the set of all services that can be produced at a cost not exceeding C. The cost indirect output distance function is defined as the maximal expansion (or the minimal contraction of the reciprocal) of services delivered that can be produced by a combination of resources which satisfies the cost constraint (see Balk 1998, p. 141 and Färe and Primont 1995):

$$ID_o(y, w/C) = \inf\{\theta > 0 : (y/\theta) \in P(x), wx \leq C\} \qquad (1)$$

where ID_o = cost indirect output distance; y = vector of services delivered; w = vector of resource prices;

x = vector of resources; C = cost; $P(x)$ = production possibility set; θ is a scalar.

It should be noted that inefficiency derived from the indirect output distance is also due to a non-optimal allocation of resources. The inefficiency thus includes technical and allocative inefficiency.

The optimal resource demand equations corresponding to the cost indirect output distance function are derived from Roy's identity (see Färe and Primont 1995, p. 92):

$$S_j = \frac{w_j x_j}{C} = \frac{\partial \ln ID_o\left(y, \frac{w}{C}\right)}{\partial \ln\left(\frac{w_j}{C}\right)} \bigg/ \sum_n \left[\frac{\partial \ln ID_o\left(y, \frac{w}{C}\right)}{\partial \ln\left(\frac{w_n}{C}\right)}\right] \quad (2)$$

where S_j = cost share resource j ($j = 1,...,N$); w_j = jth resource price ($j = 1,...,N$).

It should be noted that only under non-constant returns to scale technologies the indirect output distance approach has an added value compared to a cost function approach. As shown by Färe and Primont (1995, p. 83), the indirect output distance function equals the cost function if and only if the technology exhibits constant returns to scale.

Along the lines of the original shadow cost function approach, as discussed by Kumbhakar (1997), I assume that w^* is a vector of shadow resource prices that corresponds to output maximizing resource quantities x. Note that here it is assumed that actual resource quantities correspond with the optimal values. Shadow cost C^* is the corresponding cost at shadow resource prices w^*. Therefore, the vector of shadow resource prices w^* is the solution to the following system of equations:

$$ID_o\left(y, \frac{w^*}{C^*}\right) = 1 \quad (3)$$

$$S_j^* = s_j\left(y, \frac{w^*}{C^*}\right) \quad (4)$$

where, S_j^* = shadow cost share resource j ($j = 1,...,N$); w_j^* = shadow resource price j ($j = 1,...,N$); C^* = shadow cost; and $s_j(.)$ ($j = 1,...,N$) are the corresponding equations derived from Roy's identity (see Eq. 2). Shadow cost and shadow cost shares are deducted from actual resource quantities and shadow resource prices.

It is assumed that the cost indirect output distance function is a continuous twice differentiable function that satisfies all the duality requirements, such as non-decreasing and quasi concave in w^*/C^* and non-decreasing, convex and homogeneous of degree one in y. It is further assumed that (firm-specific) shadow resource prices are proportional to actual resource prices:

$$w_j^* = k_j \cdot w_j \quad (5)$$

where, k_j = distortion factor of resource j.

Note that k_j can vary between firms. For reasons of simplicity, we do not denote the firm by a separate index. Since we assume that the vector of actual resource quantities x are the optimal quantities at shadow resource price w^*, we calculate the algebraic relationship between actual and shadow cost as follows:

$$C^* = \sum_j w_j^* x_j = \sum_j k_j w_j x_j\left(y, \frac{w^*}{C^*}\right) = \sum_j k_j \frac{w_j x_j\left(y, \frac{w^*}{C^*}\right)}{C} \cdot C$$

$$= C \cdot \sum_j k_j S_j \quad (6)$$

For computational ease we define:

$$G_j = \left(\frac{k_j}{\sum_j k_j S_j}\right) \quad (7)$$

For reasons of convenience, we also derive an expression for $\ln\left(\frac{w_j^*}{C^*}\right)$ by substituting Eqs. 5 and 7:

$$\ln\left(\frac{w_j^*}{C^*}\right) = \ln\left(\frac{w_j}{C}\right) + \ln(k_j) - \ln\left(\sum_j k_j S_j\right)$$

$$= \ln\left(\frac{w_j}{C}\right) + \ln(G_j) \quad (8)$$

Substituting shadow resource prices and shadow cost into Eq. 2 yields:

$$S_j^* = \frac{\partial \ln ID_o\left(y, \frac{w^*}{C^*}\right)}{\partial \ln\left(\frac{w_j^*}{C^*}\right)} \bigg/ \sum_n \frac{\partial \ln ID_o\left(y, \frac{w^*}{C^*}\right)}{\partial \ln\left(\frac{w_n^*}{C^*}\right)} \cdot \quad (9)$$

From this, we derive the actual cost shares by substituting (5) in (9) and using the expression (7):

$$S_j = \frac{w_j x_j}{C} = \frac{\left(\frac{w_j}{w_j^*}\right) \cdot w_j^* \cdot x_j}{C^* \cdot \left(\frac{1}{\sum_j k_j S_j}\right)} = \left(\frac{\sum_j k_j S_j}{k_j}\right) \cdot S_j^* = \frac{S_j^*}{G_j} \quad (10)$$

Equations 4 and 10 form an alternative system of the cost indirect output distance function based on shadow resource prices and shadow cost. In Sect. 3, we will derive a complete system of equations from a translog specification.

3 Translog function

One of the most popular functions in analyzing cost and output distance functions is the translog function (see Christensen et al. 1973). The translog function consists of single and second order terms of all variables in logarithms. In the case of the SCIOD this yields:

$$\ln ID_o\left(y,\frac{w^*}{C^*}\right) = \left\{ a_0 + \sum_m b_m \ln(y_m) + \frac{1}{2}\sum_m \sum_{m'} b_{mm'} \right.$$
$$\times \ln(y_m)\ln(y_{m'}) + \sum_n c_n \ln\left(\frac{w_n^*}{C^*}\right)$$
$$+ \frac{1}{2}\sum_n \sum_{n'} c_{nn'} \ln\left(\frac{w_n^*}{C^*}\right)\ln\left(\frac{w_{n'}^*}{C^*}\right)$$
$$\left. + \sum_m \sum_n e_{mn} \ln(y_m)\ln\left(\frac{w_n^*}{C^*}\right) \right\}$$
(11)

Substituting Eqs. 5, 6, and 8 into Eq. 11 yields:

$$0 = \ln TL\left(y,\frac{w}{C}\right) + \ln AE \tag{12}$$

with:

$$\ln TL\left(y,\frac{w}{C}\right) = \left\{ a_0 + \sum_m b_m \ln(y_m) + \frac{1}{2}\sum_m \sum_{m'} b_{mm'} \right.$$
$$\times \ln(y_m)\ln(y_{m'}) + \sum_n c_n \ln\left(\frac{w_n}{C}\right)$$
$$+ \frac{1}{2}\sum_n \sum_{n'} c_{nn'} \ln\left(\frac{w_n}{C}\right)\ln\left(\frac{w_{n'}}{C}\right)$$
$$\left. + \sum_m \sum_n e_{mm} \ln(y_m)\ln\left(\frac{w_n}{C}\right) \right\} \tag{13}$$

and:

$$\ln AE = \sum_n c_n \ln(G_n) + \frac{1}{2}\sum_n \sum_{n'} c_{nn'} \ln(G_n)\ln(G_{n'})$$
$$+ \frac{1}{2}\sum_n \sum_{n'} c_{nn'} \ln\left(\frac{w_n}{C}\right)\ln(G_{n'})$$
$$+ \frac{1}{2}\sum_n \sum_{n'} c_{nn'} \ln(G_n)\ln\left(\frac{w_{n'}}{C}\right)$$
$$+ \sum_m \sum_n e_{mn} \ln(y_m)\ln(G_n) \tag{14}$$

where, $TL(.)$ = translog function; $\ln AE$ = allocative efficiency component.

Note that at the frontier $ID_o(y, w^*/C^*) = 1$ and that the ratios of shadow resource prices and shadow cost (w_n^*/C^*) in (11) are substituted by the ratio of actual resource prices and actual cost (w_n/C) and G_n. The terms which include G_n are consolidated in $\ln AE$ (14). $\ln AE$ represents the percentage of potential increase in services delivered due to allocative inefficiency. Allocative efficiency thus depends on the distortion factors k_j, actual cost shares and services delivered.

Equation 11 has now analytically been separated in two components. One part consists of the relationship between actual services, resource prices and cost and the second part refers to the impact of using shadow resource prices

and shadow cost. The first part reflects the frontier; the second part reflects the deviations from the optimal resources at actual resource prices (allocative inefficiency). A possible third part may reflect the distance to the frontier by radial expansion (technical efficiency component) and can be added to Eq. 12 and be derived econometrically (see Sect. 4).

By applying (10) to (11), cost share equations can be derived. For resource j the cost share equals:

$$S_j = \left(\frac{S_j^*}{G_j}\right)$$
$$= \left(\frac{1}{G_j}\right)$$
$$\times \frac{c_j + \sum_n c_{jn}\left[\ln\left(\frac{w_n}{C}\right) + \ln(G_n)\right] + \sum_m e_{mj}\ln(y_m)}{\sum_k \left[c_k + \sum_n c_{kn}\left[\ln\left(\frac{w_n}{C}\right) + \ln(G_n)\right] + \sum_m e_{mk}\ln(y_m)\right]} \tag{15}$$

where, y_m = service m ($m = 1,\ldots,M$); w_n = resource price n ($n = 1, \ldots ,N$); $a_0, b_m, b_{mm'}, c_n, c_{nn'}, e_{mn}$ parameters to be estimated.

A number of additional restrictions are required before we can estimate this model. Since the translog function is continuous and twice differentiable with respect to y and w, symmetry automatically holds;

$$b_{mm'} = b_{m'm}; \quad c_{nn'} = c_{n'n}$$

Since the SCIOD is linear homogenous in y, the following restrictions also have to be imposed on the parameters:

$$\sum_m b_m = 1; \qquad \sum_m b_{mm'} = 0 \ (m' = 1,\ldots,M);$$
$$\sum_m e_{mn} = 0 \ (n = 1,\ldots,N)$$

Economies of scale can be derived from the output ray elasticity, defined by (see Färe and Zieschang 1991):

$$roe(.) = \sum_n \frac{\partial \ln ID_o(.)}{\partial \ln(w_n/C)} \tag{16}$$

where, $roe(.)$ = function for ray output elasticity.

Marginal cost of service y_m equals (see Färe and Zieschang 1991):

$$mc(y_m) = \left(\frac{C}{roe(.)}\right) \cdot \frac{\partial \ln ID_o(.)}{\partial \ln(y_m)} \tag{17}$$

where, $mc(y_m)$ = function for marginal cost service m.

4 Estimation

One of the advantages of the proposed approach is that it removes the allocative efficiency component from the error term of the cost share equations. However, estimation of

the above system is far from straightforward. There is a trade-off with the many non-linearities in the model through the distortion factors k_{ij}. Since the allocative component in the model is modeled explicitly, allocative inefficiency can be derived from the estimated parameters and the observed values for services delivered, resource prices and cost. A key element in this Eq. 14 is G_j depending on the distortion factor k_j. The parameter k_j may reflect firm-specific input preferences, but also simply systematic preferences. For firm-specific preferences, k_j (or better k_{ij}) varies between firms and can only be estimated by using panel data or by assuming that k_j depends on other firm specific managerial characteristics (see e.g., Reinhard and Thijssen 2000). For systematic preferences, k_j is constant across firms and can be estimated simply. The technical efficiency component may be regarded as a part of the error term and can be derived from the residuals. Depending on the structure of the available data, a number of methods are available (see e.g., Kumbhakar and Lovell 2000). For instance, in the case of panel data and a limited number of firms, a fixed-effect approach is one possible way to proceed. However, there are a number of alternative specifications of the error term.

Since we are only interested in the effect of relative resource prices, one of the distortion factors has to be set to one. Another possibility is to normalize the distortion factors in such way that shadow cost equals actual cost. At first sight, this seems to be an attractive way to proceed (see Balk 1997), but, as Maietta (2002) points out, outcomes become sensitive to the choice of the input price as numéraire. Nevertheless, both approaches are equivalent (see Kumbhakar and Karagiannis 2004); a trade-off needs to be made between the simplicity of the model and the sensitivity of the numéraire choice, as mentioned. The same problem occurs in applying the SCIOD.

In traditional cost equation system estimation, cost shares only appear as endogenous variables on the left-hand side of the cost-share equations and exogenous variables as service levels and resource prices on the right-hand side. Estimation is carried out in a relatively straightforward way by applying Zellner's Seemingly Unrelated Regression (Zellner 1962). Since cost shares here are endogenous variables that appear on both sides (note S_j's are included in G_j's) of the Eqs. 14 and 15 in the system, it is no longer appropriate to apply SUR. The system should therefore be estimated by using minimum distance estimation techniques, such as the General Method of Moments, Full Information Maximum Likelihood and Iterative Nonlinear 3SLS. In general, these techniques are much more demanding from a computational point of view and can only be applied under the assumption of the same distortion factor for all firms. Some of these techniques may even fail to converge due to problematic starting

values and so forth. Due to the singularity of the model, one of the share equations should be dropped from the model. An alternative approach is based on Bayesian inference. For a detailed discussion of a Bayesian application to a shadow cost function, I recommend reading Kumbhakar and Tsionas (2005). Another alternative is based on an iterative procedure. This procedure consecutively consists of setting the distortion factors at a fixed value, estimating the other parameters of the model and simulating the optimal values for the distortion parameters. This procedure is repeated until some convergence criterion is met. Furthermore, it is obvious that the estimation of shadow prices depends to a large extent on sufficient variation in resource prices, which may not be the case in many empirical applications.

5 Conclusions

In many public service industries, firms are constrained by a cost budget and characterized by non-maximizing output behavior. None of the popular empirical models described in literature, such as the cost function model, is suited to analyzing the production structure of these types of service industries. In this paper, an alternative model, referred to as a Shadow Cost Indirect Output Distance Function, is presented that is more suitable. The model requires the same data as a cost function model. It provides an explicit expression for allocative efficiency that also depends on other exogenous variables in the model. Estimation of the system, however, is rather difficult to conduct. Although some promising techniques are at hand, much more empirical experience is needed in this area.

References

Atkinson SE, Primont D (2002) Stochastic estimation of firm technology, inefficiency, and productivity growth using shadow cost and distance functions. J Econom 108:203–225

Atkinson SE, Cornwell C (1994) Estimation of output and input technical efficiency using a flexible functional form and panel data. Int Econ Rev 35(1):245–255

Balk BM (1997) The decomposition of cost efficiency and the canonical form of cost function and cost share equations. Econ Lett 55:45–51

Balk BM (1998) Industrial prices, quantity, and productivity indices: the micro-economic theory and an application. Kluwer Academic Publishers, Dordrecht

Blank JLT, Merkies AHQM (2004) Empirical assessment of the economic behaviour of Dutch general hospitals. Health Econ 13:265–280

Blank JLT (2000) Public provision and performance: contributions from productivity and efficiency measurement. Elsevier, Amsterdam

Christensen LR, Jorgenson DW, Lau LJ (1973) Transcendental logarithmic production frontiers. Rev Econ Stat 55(1):28–45

Dor A, Duffy S, Wong H (1997) Expense preference behavior and contract-management: evidence from U.S. hospitals. South Econ J 64(2):542–554

Eakin BK (1993) Do physicians minimize cost? In: Fried HO, Lovell CAK, Schmidt SS (eds) The measurement of productive efficiency: techniques and applications. Oxford University Press, Oxford, pp 221–236

Eakin BK, Kniesner TJ (1988) Estimating a non-minimum cost function for hospitals. South Econ J 54(3):583–592

Färe R, Zieschang KD (1991) Determining output shadow prices for a cost-constrained technology. J Econ 54(2):143–155

Färe R, Primont D (1995) Multi-output production and duality: theory and applications. Kluwer Academic Publishers, Boston

Grosskopf S, Haynes KJ, Taylor LL, Weber WL (1997) Budget-constrained frontier measures of fiscal equality and efficiency in schooling. Rev Econ Stat 79(1):116–124

Hayes KJ, Razzolini L, Ross LB (1998) Bureaucratic choice and non-optimal provision of public goods: theory and evidence. Public Choice 94:1–20

Kumbhakar SB (1997) Modeling allocative inefficiency in a translog cost function and cost share equations: an exact relationship. J Econom 76:351–356

Kumbhakar SC, Lovell CAK (2000) Stochastic frontier analysis. Cambridge University Press, New York

Kumbhakar SB, Karagiannis G (2004) On the equivalence of two normalizations in estimating shadow cost functions. Econ Lett 82:15–19

Kumbhakar SC, Tsionas EG (2005) The joint measurement of technical and allocative inefficiencies: an application of Bayesian inference in nonlinear random-effects models. J Am Stat Assoc 100:736–747

Maietta OW (2000) The decomposition of cost efficiency into the technical and allocative components with panel data of Italian dairy farms. Eur Rev Agric Econ 27(4):473–495

Maietta OW (2002) The effect of the normalization of the shadow price vector on the cost function estimation. Econ Lett 77:381–385

Reinhard S, Thijssen G (2000) Nitrogen efficiency of Dutch dairy farms: a shadow cost function approach. Eur Rev Agric Econ 27(2):167–186

Rodrigues-Alvarez A, Lovell CAK (2004) Excess capacity and expense preference behaviour in National Health Systems: an application to the Spanish public hospitals. Health Econ 13:157–169

Zellner A (1962) An efficient method of estimating seemingly unrelated regression and tests of aggregation bias. J Am Stat Assoc 57:500–509