Applied Cost Allocation

The DEA–aumann–shapley Approach

Bogetoft, Peter; Hougaard, Jens Leth; Smilgins, Aleksandrs

Document Version
Accepted author manuscript

Published in:
European Journal of Operational Research

DOI:
10.1016/j.ejor.2016.04.023

Publication date:
2016

Applied Cost Allocation: The DEA-Aumann-Shapley Approach

Peter Bogetoft, Jens Leth Hougaard, and Aleksandrs Smilgins

Journal article (Post print version)

CITE: Applied Cost Allocation : The DEA-Aumann-Shapley Approach. / Bogetoft, Peter; Hougaard, Jens Leth; Smilgins, Aleksandrs. In: European Journal of Operational Research, Vol. 254, No. 2, 2016, p. 667-678.

DOI: http://dx.doi.org/10.1016/j.ejor.2016.04.023

Applied Cost Allocation: The DEA-Aumann-Shapley Approach

Peter Bogetoft

Department of Economics

Copenhagen Business School

Jens Leth Hougaard & Aleksandrs Smilgins

Department of Food and Resource Economics

University of Copenhagen

February 2016

Abstract

This paper deals with empirical computation of Aumann-Shapley cost shares for joint production. We show that if one uses a mathematical programming approach with its non-parametric estimation of the cost function there may be observations in the data set for which we have multiple Aumann-Shapley prices. We suggest to overcome such problems by using lexicographic goal programming techniques. Moreover, cost allocation based on the cost function is unable to account for differences between efficient and actual cost. We suggest to employ the notion of *rational inefficiency* in order to supply a set of assumptions concerning firm behavior. These assumptions enable us to connect inefficient with efficient production and thereby provide consistent ways of allocating the costs arising from inefficiency.

Keywords: Cost Allocation, Convex Envelopment, Data Envelopment Analysis, Aumann-Shapley Pricing, Inefficient Joint Production.

JEL classification: C61, D24, C13, C63, D61.

Correspondence: Jens Leth Hougaard, Department of Food and Resource Economics, University of Copenhagen, Rolighedsvej 25, 1958 Frederiksberg C., Denmark.
E-mail: jlh@ifro.ku.dk

Acknowledgements: The authors are grateful for helpful comments from Rick Antle, Mette Asmild, Emma Potter, and, Jørgen Tind.

1 Introduction

Aumann-Shapley (A-S) cost allocation, often interpreted as generalized average cost sharing, is a well-known cost allocation method designed for regulation of multi-product natural monopolies as well as for internal cost accounting and decentralized decision making in organizations, see e.g., Spulber (1989), Banker (1999), Mirman, Tauman and Zang (1985a). In short, the idea is to determine a set of unit prices for each output, i.e., the Aumann-Shapley (A-S) prices, and use these for the allocation of joint costs.

The theoretical literature has shown that the A-S method (and A-S prices) possesses a number of desirable properties, see e.g., Billera and Heath (1982), Mirman and Tauman (1982), Young (1985), Mirman, Tauman, and Zang (1985b), and it has essentially been the unanimous recommendation of economists for decades when sharing the costs of joint production, see e.g., Friedman and Moulin (1999). Yet, despite its sound theoretical foundation there has been relatively few empirical applications. The reason seems at least twofold:

1. It requires an empirical estimation of the cost function that enables computation of all relevant A-S prices.

2. In practice firms may not produce at efficient production cost. Hence, an allocation based solely on the cost function will not account for differences between efficient and actual costs.

In the present paper we examine how to cope with both these issues. While there has been previous papers dealing with computation of A-S prices for given empirical cost functions we believe the second issue, concerning inefficient production, has been ignored and we offer a completely new approach here.

We follow up on papers by Samet, Tauman and Zang (1984), and Hougaard and Tind (2009) and consider empirical estimation based on convex envelopment of observed cost-output data as in the celebrated Data Envelopment Analysis (DEA) approach of Charnes, Cooper and Rhodes (1978).[1] The resulting piecewise linear cost function enables a relatively simple computation of A-S prices for large parts of the output space: The A-S prices associated with a given output vector are simply found as the weighted sum of gradients of the linear facets of the estimated cost function along a radial contraction path of the observed output vector, where the weights are proportional to the length of the projected line segments. For every data point this can be computed using parametric linear programming.

[1]For a recent general DEA reference, see e.g., Bogetoft and Otto (2010).

However, for certain data points, and in particular for the observed productions that help span the empirical cost function, the cost function will in most cases be piecewise continuous differentiable along the radial contraction path and hence there may be multiple A-S prices for the same observation caused by lack of continuous differentiability on subintervals along this path. For reasons of transparency and simplicity, we suggest to overcome this problem by using a lexicographic goal programming approach with a predefined ordering of outputs to determine which outputs should be allocated most costs. Such orderings may, for instance, be the result of a managerial prioritization and will provide unique A-S prices for all observations. [2]

Our approach, however, does not exclude the possibility of having zero A-S prices for some units (when these are referring to exterior facets). To solve this problem (as well as excluding the possibility of infinite A-S prices), one can use the "extended facet" approach in Olesen and Petersen (1996, 2003). This ensures well defined rates of substitution on the boundary of the convex envelopment of the data points, but in general this approach lacks operationability.

When it comes to inefficient production it seems that no previous papers have considered the consequences in relation to A-S pricing. Yet, countless empirical studies have shown that observed production data are often associated with considerable levels of technical inefficiency, see e.g., Bogetoft and Otto (2010).

To deal with inefficient production in the context of A-S cost allocation, we propose to invoke the *rational inefficiency paradigm* introduced in Bogetoft and Hougaard (2003) and further analyzed in Asmild, Bogetoft and Hougaard (2009, 2013). This allows us to formulate specific assumptions concerning the behavior of inefficient firms, which in turn enables us to associate an efficient production with each inefficient observation in the sample. It is worth emphasizing that, as such, our suggested approach and associated results are independent of the way we estimate the cost function (although we are using non-parametric estimation for our empirical illustration).

In particular, firms can introduce inefficiency on either the cost (input) side or the production (output) side. Considering cost inefficiency we assume that the inefficient firm has revealed a constant fraction of overspending by its actual production choice. Thus, A-S prices connected with the cost efficient production can be scaled up with a radial cost efficiency index in order to obtain full cost allocation. We show that this approach is tantamount to viewing inefficiency as a fixed cost and to sharing this fixed cost in proportion to the A-S prices.

[2] A far less operational approach would be to determine all facets involved (for data point in question) and define the associated A-S price as the (weighted) average of the gradients of these facets.

Looking at the output side we assume that firms introduce inefficiency by consuming outputs directly on the job. For example, some units of a given output may be produced in inferior quality and we can regard this as a kind of internal ”consumption”, which should not distort the estimation of A-S prices. A rationally inefficient firm would choose its actual (unobserved) production so as to maximize potential revenue given output prices and its observed cost. When we observe the actual output level lower than that it is because the firm has consumed the difference (slack) itself. We shall therefore argue that it is the allocatively efficient output combination that carries the cost and allocate costs accordingly using the A-S prices related to the allocatively efficient production.

We illustrate our approach using a data set concerning Danish waterworks. We use the same 2011 data that the regulator, the Water Division of the Danish Competition and Consumer Authority, used in their first regulatory cost benchmarking model, and we show how cost shares can be computed using our suggested A-S approach in case of a non-parametric estimation of the cost function.

The rest of the paper is organized as follows: Section 2 defines the standard Aumann-Shapley cost allocation rule for continuously differentiable cost functions. Section 3 introduces the convex envelopment approach to the estimation of the empirical cost function. We discuss how to calculate A-S prices from the estimated cost function and suggest how to deal with the lack of well defined A-S prices for all production units in section 4. Section 5 deals with inefficient production in the context of A-S cost allocation building on the rational inefficiency paradigm. The illustrative application to data on Danish waterworks is presented in section 6, and section 7 contains final remarks.

2 Aumann-Shapley Cost Allocation

Consider a joint production process resulting in n different outputs. Let $q \in \mathbf{R}^n_+$ be the (non-negative) output vector where q_i is the level of output i. The cost of producing any vector q is given by a non-decreasing cost function $C : \mathbf{R}^n_+ \to \mathbf{R}$. Initially, we assume that $C(0) = 0$, i.e., there are no fixed costs.

Let (q,C) denote a cost allocation problem and let ϕ be a cost allocation rule. The cost allocation rule specifies a unique vector of cost shares $x = (x_1,\ldots,x_n) = \phi(q,C)$ for each output vector q and cost function C. The cost shares satisfy budget-balance, i.e.

$$\sum_{i=1}^{n} x_i = C(q)$$

where x_i is the cost share allocated to output i.

In particular, consider the class of continuously differentiable cost functions and let $\partial_i C(q) = \partial C(q)/\partial q_i$ be the partial derivative of C at q with respect to the ith argument.

Following Aumann and Shapley (1974), we define the Aumann-Shapley rule (A-S-rule) ϕ^{AS} as

$$\phi_i^{AS}(q,C) = \int_0^{q_i} \partial_i C(\frac{t}{q_i}q)dt = q_i \int_0^1 \partial_i C(tq)dt \text{ for all } i = 1,\dots,n. \tag{1}$$

It can be shown that this allocation is budget balanced, i.e., $\sum_{i \in N} \phi_i^{AS}(q,C) = C(q)$.

Also,

$$p_i^{AS} = \int_0^1 \partial_i C(tq)dt$$

can be seen as the unit cost of output i. This is known as the Aumann-Shapley price (A-S price) of output i. As such, the A-S cost shares, x_i^{AS}, are given by

$$x_i^{AS} = p_i^{AS} q_i \tag{2}$$

for all outputs $i = 1,\dots,n$.

The A-S rule can be seen as one (of several) possible extensions of average cost sharing to the multiple product case, see e.g. Hougaard (2009). Axiomatic characterizations are provided (independently) in Billera and Heath (1982) and Mirman and Tauman (1982). Following the latter, we here shortly recall the axioms characterizing A-S pricing $p^{AS}(C,q)$:

- (Rescaling) For some rescaling $q \mapsto \bar{q} = (\lambda_1 q_1,\dots,\lambda_n q_n)$, let $G(q) = C(\bar{q})$. Then, for all $i = 1,\dots,n$, $p_i(G,q) = \lambda_i p_i(C,\bar{q})$.

- (Consistency) Let $C(q) = G(\sum_{i=1}^n q_i)$. Then, for all $i = 1,\dots,n$, $p_i(C,q) = p_i(G,\sum_{i=1}^n q_i)$.

- (Additivity) Let $C(q) = G(q) + H(q)$. Then $p(C,q) = p(G,q) + p(H,q)$.

- (Positivity) Let C be non-decreasing at each $q' \leq q$. Then $p(C,q) \geq 0$.

Early examples of application can be found in Billera, Heath and Raanan (1978) and Samet, Tauman and Zang (1983). More recent applications include Castano-Pardo and Garcia-Diaz (1995), Haviv (2001), Tsanakas and Barnett (2003) and Bjørndal and Jörnsten (2005).

Example 1: Consider the simple case where the cost function is homogeneous of degree k, i.e., $C(tq) = t^k C(q)$ for $t \in [0,1]$. Here it is clear that for all $i \in N$, the A-S prices become

$$p_i^{AS} = \partial_i C(q) \int_0^1 t^{k-1} dt = p_i^{MC} \frac{1}{k},$$

and hence,

$$\phi_i^{AS}(q,C) = q_i p_i^{MC} \frac{1}{k},$$

where p_i^{MC} is the marginal cost price of i. $\triangle$

If the production involves fixed costs, $C(0) > 0$, the above procedure will not allocate the full cost, only the variable part, $C(q) - C(0)$. An obvious idea in this case is to allocate the fixed part, $C(0)$, in proportion to the variable shares, i.e., to let $\widetilde{C}(q) = C(q) - C(0)$ and use

$$\phi_i^{AS}(q,C) = \phi_i^{AS}(q,\widetilde{C}) + \frac{\phi_i^{AS}(q,\widetilde{C})}{\sum_{j=1}^n \phi_j^{AS}(q,\widetilde{C})} C(0), \tag{3}$$

for all $i = 1, \ldots, n$.

Mirman, Samet and Tauman (1983) presents an axiomatic characterization of an (extended) A-S pricing rule for cost functions with a fixed cost component corresponding to the extended A-S cost allocation rule (3). It turns out that a slight modification of the original axioms as stated above, and in particular, a modified version of additivity where it is claimed that there exists a way to split the fixed cost between the different variable cost components, is enough to achieve this:

- (Modified Additivity) For each $G \leq \widetilde{C}$ there is a non-negative number C_G^0 such that if $\widetilde{C} = \sum_{i=1}^m G_i$ then $C(0) = \sum_{i=1}^m C_{G_i}^0$ and

  $$p(\widetilde{C} + C(0), q) = \sum_{i=1}^m p(G_i + C_{G_i}^0, q).$$

  Moreover, $G_i(q) \geq G_j(q)$ implies $C_{G_i}^0 \geq C_{G_j}^0$.

3 Non-Parametric Cost Functions and Efficiency Measures

To apply the A-S rule in practice, the cost function C needs to be estimated. Let $\mathscr{D} = \{(q_j, C_j)\}_{j=1,\ldots,h}$ be a set of h observations of output vectors $q_j \in \mathbf{R}_+^n$ and their associated production cost $C_j \in \mathbf{R}_+$. These observations can be construed as originating either from the same firm over h time periods or from h individual firms at a given point in time.

Based on such a data set, a cost function may be estimated using a traditional parametric approach where a given functional form is postulated and parameters associated with this form are estimated. This approach is taken in numerous studies of cost functions, see e.g., Kumbhakar and Lovell (2000). However, using a standard parametric approach (typically estimating parameters in log-linear or translog functional forms) may produce significantly biased estimates and provide invalid inference as shown in Delis, Iosifidi and Tsionas (2014), who argue that a semi-parametric smooth coefficient model yields better results. Indeed, to use various approaches for smoothing represents one way of obtaining well-defined A-S prices within a conventional econometric framework.

Alternatively, one may take a (non-smooth) non-parametric, mathematical programming approach. Since this approach is widely used for regulation of various forms of natural monopolies (see e.g., Bogetoft, 2012) we will focus on what is essentially the so called Data Envelopment Analysis (DEA) approach, first introduced by Charnes, Cooper and Rhodes (1978) and since then extended and applied in thousands of Operations Research papers. In the context of cost functions, we can think of the approach as follows: We assume that the underlying but unknown cost function $C(\cdot)$ have the following properties

$A1$	Increasing: $q' \geq q \Rightarrow C(q') \geq C(q)$
$A2$	Convex: $C(\alpha q + (1-\alpha)q') \leq \alpha C(q) + (1-\alpha)C(q'), \forall \alpha \in [0,1]$
$A3(\gamma)$	$\gamma -$ returns to scale: $C(\kappa q) \leq \kappa C(q), \ \ \forall \kappa \in \Gamma(\gamma)$

where γ represents different global returns to scale properties, namely, varying vrs, decreasing drs, increasing irs, and constant crs returns to scale, with corresponding parameter sets $\Gamma(vrs) = \{1\}$, $\Gamma(drs) = [0,1]$, $\Gamma(irs) = [1,\infty)$, and $\Gamma(crs) = [0,\infty)$.

It follows that the initial uncertainty about how to represent the cost function can be expressed as uncertainty within the following broad classes of cost functions:

$$\begin{aligned}
\mathscr{C}(\text{crs}) &= \{C : \mathbf{R}^n_+ \to \mathbf{R}_+ \mid C \text{ is increasing, convex, crs}\} \\
\mathscr{C}(\text{drs}) &= \{C : \mathbf{R}^n_+ \to \mathbf{R}_+ \mid C \text{ is increasing, convex, drs}\} \\
\mathscr{C}(\text{irs}) &= \{C : \mathbf{R}^n_+ \to \mathbf{R}_+ \mid C \text{ is increasing, convex, irs}\} \\
\mathscr{C}(\text{vrs}) &= \{C : \mathbf{R}^n_+ \to \mathbf{R}_+ \mid C \text{ is increasing, convex, vrs}\}
\end{aligned}$$

Thus, the idea of the non-parametric approach is that there is considerable a priori uncertainty about the functional form of the cost function. In a γ-model, all we know is that

$$C(\cdot) \in \mathscr{C}(\gamma)$$

i.e., we know that $C(\cdot)$ is increasing and convex, but otherwise we know nothing about the cost function. Our a priori uncertainty does not stem from a lack of information about a few parameters, as in a Coob-Douglas or a Translog statistical model. Rather, we lack information about all of the characteristics of the function except for a few general properties such as its convexity and tendency to increase.

Now, to estimate a specific cost function from the available data $\mathscr{D} = \{(q_j, C_j)\}_{j=1,\ldots,h}$ we can rely on the the *minimal extrapolation principle*. We estimate the non-parametric cost function as the largest function with these properties that is consistent with data in the sense that $C_j \geq C(q_j)$, $j = 1,\ldots,h$, i.e., as

$$\widehat{C}(q) := \max\{C(q) \mid C(\cdot) \in \mathscr{C}(\gamma), C_j \geq C(q_j), \quad j = 1,\ldots,h\}$$

An important feature of the class of cost functions considered above is that a largest function of this type actually exists – or, to put it differently, that the $\widehat{C}(\cdot)$ defined above inherits the properties from the $\mathscr{C}(\gamma)$ class. This is shown for example in Bogetoft (1997).

It is easy to show that the estimated cost function in the drs case is the solution to the following program

$$\begin{aligned}
\widehat{C}^{drs}(q^*) = \min \quad & \theta & (4)\\
\text{s.t.} \quad & \sum_{j=1}^{h} \lambda_j q_j \geq q^* & (5)\\
& \sum_{j=1}^{h} \lambda_j C_j \leq \theta & (6)\\
& \sum_{j=1}^{h} \lambda_j \leq 1 & (7)\\
& \lambda_j \geq 0, \forall j. & (8)
\end{aligned}$$

For the other classes of cost functions we get similar mathematical programs; the only difference is that the constraint (7) is modified. It is $\sum_{j=1}^{h} \lambda_j = 1$ in the vrs case, $\sum_{j=1}^{h} \lambda_j \geq 1$ in the irs case, and it is ignored in the crs case. We say that an observation (q^*, C^*) is *efficient* if $C^* = \widehat{C}^{\gamma}(q^*)$.

It follows from this expression of the estimated cost function that it will be convex and piecewise linear.[3] Consequently, it is clear that the estimated cost function will not be

[3]The convexity assumption may be relaxed (see, e.g., Bogetoft 1996, Bogetoft, Tama and Tind 2000) but, for the present purpose and the ease of exposition, we stick to convexity.

differentiable in general. In fact, local non-differentiability of $\widehat{C}(q)$ may create problems when determining the A-S prices (and thereby the A-S cost shares) for specific data points. A problem which we return to in the next section.

Since we will also be concerned with cost allocation in the presence of inefficiency, we further introduce a few efficiency measures. In the spirit of Debreu (1951) and Farrell (1957) we can measure the efficiency of the observation (q^*, C^*) relative to the (estimated) cost function $\widehat{C}$ using the radial indexes,

$$E^q((q^*, C^*), \widehat{C}) = \max\{\theta \in \mathbf{R} \mid C^* \geq \widehat{C}(\theta q^*)\} \in [1, \infty) \tag{9}$$

and

$$E^c((q^*, C^*), \widehat{C}) = \min\{\theta \in \mathbf{R} \mid \theta C^* \geq \widehat{C}(q^*)\} \in [0, 1] \tag{10}$$

By (9) we get that, given the cost level C^*, the firm could have produced output level $E^q q^*$ if it was efficient. By (10) we get that, given the output production q^*, the firm could have produced at costs $E^c C^*$ if it was efficient. Under crs it is well known that[4],

$$E^c((q^*, C^*), \widehat{C}^{crs}) = \frac{1}{E^q((q^*, C^*), \widehat{C}^{crs})} \tag{11}$$

4 Estimation of Aumann-Shapley Cost Shares

Following the approach in Hougaard and Tind (2009), calculation of A-S cost shares with respect to an estimated cost function $\widehat{C}$ can be done using parametric programming.[5] Indeed, the A-S prices (and hence cost shares) are easily determined as a finite sum of gradients of the linear pieces of $\widehat{C}$ along the line segment $[0, q]$ weighted with the normalized length of the subintervals where $\widehat{C}$ has constant gradient.

Specifically, consider a given output vector q^* and a parameter value $t \in [0, 1]$. Now, we

[4] see e.g., Bogetoft and Otto (2010).

[5] See, e.g., Bazaraa, Jarvis and Sherali (1990).

can solve the following linear programming problem

$$\min \quad \sum_{j=1}^{h} \lambda_j C_j \tag{12}$$

$$\text{s.t.} \quad \sum_{j=1}^{h} \lambda_j q_j \geq t q^* \tag{13}$$

$$\sum_{j=1}^{h} \lambda_j \in \Gamma(\gamma) \tag{14}$$

$$\lambda_j \geq 0, \forall j. \tag{15}$$

for all values of the parameter $t \in [0,1]$ and γ representing different global returns to scale properties with $\Gamma(vrs) = \{1\}$, $\Gamma(drs) = [0,1]$, $\Gamma(irs) = [1,\infty)$, and $\Gamma(crs) = [0,\infty)$. Note that in the crs case (14) and (15) coincide. The dual problem of (12) - (15) is given by

$$\max \quad \sum_{i=1}^{n} u_i(tq_i^*) + u_0 \tag{16}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} u_i q_i - vC + u_0 \geq 0 \tag{17}$$

$$\sum_{i=1}^{n} u_i q_i \in \Gamma(\gamma) \tag{18}$$

$$u, v \geq 0, \; u_0 \text{ free.} \tag{19}$$

As a result we obtain the subintervals of $[0,q^*]$ for which the gradients are constant, i.e., a series of values t_m for which the gradient is constant on the interval $[t_{m-1}, t_m]$. The values of the gradients are equal to the optimal values u in (16) as illustrated in the example below.

Example 2: Consider the following data from five observed production plans and a classification of each observations efficiency according to the returns to scale assumption of the estimated cost function:

Obs.	q_1	q_2	C	Efficient drs	Efficient irs	Efficient vrs	Efficient crs
1	8	3	12	yes	no	yes	no
2	3	1	4	yes	yes	yes	yes
3	2	2	3	yes	yes	yes	yes
4	4	7	13	yes	no	yes	no
5	1	1	2	no	yes	yes	no

Because $\Gamma(crs) = \Gamma(drs) \cup \Gamma(irs)$, the efficient crs observations are efficient both in the drs and irs cases.

Now, let $q^* = (5,3)$ and define $\bar{q} = (\bar{q}_1(t), \bar{q}_2(t)) = (q_1^* t, q_2^* t) = (5t, 3t)$. Since the estimated cost function differs under the four different returns to scale assumptions we get: $\widehat{C}^{drs}(5,3) = \widehat{C}^{vrs}(5,3) = 8.25$, and $\widehat{C}^{crs}(5,3) = \widehat{C}^{irs}(5,3) = 7$.

Solving (16)-(19) above we get

t	obj. func., drs	obj. func., irs	obj. func., vrs	obj. func., crs
$0 \leq t < 0.2$	$1.25\bar{q}_1 + 0.25\bar{q}_2$	$0\bar{q}_1 + 0\bar{q}_2 + 2$	$0\bar{q}_1 + 0\bar{q}_2 + 2$	$1.25\bar{q}_1 + 0.25\bar{q}_2$
$0.2 \leq t < 0.5$	$1.25\bar{q}_1 + 0.25\bar{q}_2$	$\bar{q}_1 + 0\bar{q}_2 + 1$	$\bar{q}_1 + 0\bar{q}_2 + 1$	$1.25\bar{q}_1 + 0.25\bar{q}_2$
$0.5 \leq t < 0.77$	$1.43\bar{q}_1 + 0.43\bar{q}_2 - 0.71$	$1.25\bar{q}_1 + 0.25\bar{q}_2$	$1.43\bar{q}_1 + 0.43\bar{q}_2 - 0.71$	$1.25\bar{q}_1 + 0.25\bar{q}_2$
$0.77 \leq t \leq 1$	$1.25\bar{q}_1 + 1.5\bar{q}_2 - 2.5$	$1.25\bar{q}_1 + 0.25\bar{q}_2$	$1.25\bar{q}_1 + 1.5\bar{q}_2 - 2.5$	$1.25\bar{q}_1 + 0.25\bar{q}_2$

First, consider the drs-case. From the above table we get Aumann-Shapley cost shares

$$x_1^{AS} = 5 \times \{1.25 \times 0.5 + 1.43 \times (0.77 - 0.5) + 1.25 \times (1 - 0.77)\} = 6.49$$

$$x_2^{AS} = 3 \times \{0.25 \times 0.5 + 0.43 \times (0.77 - 0.5) + 1.5 \times (1 - 0.77)\} = 1.76$$

Observe that $x_1^{AS} + x_2^{AS} = 8.25$ which is equal to the objective function value of the above program when $t = 1$, as it should be. The third constraint of the primal problem (14) is binding and receives in this case a non-zero dual variable value which is equal to the element -2.5 in the last row of the table. Again from the last row we see that the optimal dual variable corresponding to the first element in the output vector is equal to 1.25. Multiplication of this price by the output quantity $\hat{q}_1 = 5$ gives the value of 6.25. The difference between x_1^{AS} and this value is 0.23. The similar difference corresponding to the second element of the output vector is -2.74. The two differences add to 2.5 (ignoring small rounding errors) which is equal to the value of the optimal dual variable corresponding to the convexity constraint, as it should be. In this way the dual variable for the convexity constraint is distributed on to the values of the output vector.

Second, for the irs-case the Aumann-Shapley cost shares are $(x_1^{AS}, x_2^{AS}) = (4.63, 0.38)$ which sum up to 5, while the value of the objective function is 7. The difference is due to a fixed cost of 2, which is not surprising, since for t close to zero the objective function is given by $0\bar{q}_1 + 0\bar{q}_2 + 2$ (corresponding to a flat exterior facet). For the same reason, for vrs, we get $(x_1^{AS}, x_2^{AS}) = (4.87, 1.39)$ which sum up to 6.25 while the value of the objective function at this point is 8.25 (including the fixed cost of 2). The allocation is unchanged for any positive fixed cost since the gradients (A-S prices) are not affected by changes in the level of the whole cost possibility set: thus allocation of such fixed costs is naturally done in proportion to the A-S prices as suggested below. Finally, in the crs-case there is no fixed cost, and the allocated cost shares are $(x_1^{AS}, x_2^{AS}) = (6.25, 0.75)$ summing up to 7 exactly as they should. Note that the gradients are the same for all values of t and can therefore be directly used as A-S prices. $\triangle$

In the vrs- and irs-cases, the above approaches does not work directly, as we saw in Example 2 above, since there may be fixed costs, i.e., $\widehat{C}(0) > 0$. In this case we may proceed in two steps. First, we calculate the variable part of the cost function,

$$\widetilde{C}(q) = \widehat{C}(q) - \widehat{C}(0)$$

It is easy to show that when $\widehat{C} \in \mathscr{C}^{vrs}$, the variable part $\widetilde{C} \in \mathscr{C}^{drs}$, and similarly when $\widehat{C} \in \mathscr{C}^{irs}$, the variable part $\widetilde{C} \in \mathscr{C}^{drs}$. We can therefore use the variable part to determine A-S cost shares and then, second, allocate the fixed costs in proportion to these as explained above, i.e., as

$$\phi_i^{AS}(q, \widehat{C}^{vrs}) = \phi_i^{AS}(q, \widehat{C}^{vrs} - \widehat{C}^{vrs}(0)) + \frac{\phi_i^{AS}(q, \widehat{C}^{vrs} - \widehat{C}^{vrs}(0))}{\sum_{j=1}^{n} \phi_j^{AS}(q, \widehat{C}^{vrs} - \widehat{C}^{vrs}(0))} \widehat{C}^{vrs}(0) \qquad (20)$$

for all $i = 1, \ldots, n$, and similarly for the irs-case.

Consider now again the drs- or crs-cases. Although the Hougaard and Tind (2009) approach provides A-S prices for many production possibilities q, there are several limitations of the approach. First of all, the approach only allocates the efficient cost (as do indeed any allocation method based on the cost function). Firms that are inefficient will therefore not in general get a full allocation of their actual costs. Second, for the efficient firms, A-S prices may not be a unique because of multiple dual solutions associated with the parametric linear programming problem. Moreover for some efficient firms A-S prices may be zero. We will illustrate the latter problems below and suggest a solution to multiplicity (as mentioned in the introduction the problem of zero A-S prices can in principle to solved using the ”extended facet” approach of Olesen and Petersen, 1996). The problem of inefficient production will be addressed in Section 5.

Example 3: Imagine that we only have one observation (q^*, C^*) and assume that the underlying cost function satisfy constant returns to scale. In this case, it is intuitively obvious, that we do not have enough information to allocate costs C^* onto the different outputs. An example with two outputs is illustrated in Fig. 1 below. Here $q^* = (q_1^*, q_2^*) = (1, 1)$ and the observed cost is $C^* = 1$. Using crs we can construct a cost function along the ray from the origin $(0,0,0)$ through $(1,1,1)$. The cost estimate along this ray is $\widehat{C}(t,t) = t$. For other points in the output space, the estimated costs can be derived using an additional assumption of *free disposability*, i.e., for a given cost level, say C^*, outputs can be freely discarded. This leads to iso-cost lines as illustrated by the dotted lines in Fig. 1 below for cost level $C = 1$ and $C = 2$.

Every point of the ray from $(0,0,0)$ through $(1,1,1)$ belongs to two intersecting hyperplanes and hence partial derivatives are not uniquely defined. On the other hand, for given cost levels, when $q^1 > q^2$, the partial derivatives are 1 and 0, respectively, and the A-S prices associated with the two products are consequently 1 and 0, respectively. Likewise, when $q^1 < q^2$ all costs are associated with the second output. As such, even a slight perturbation of any ray-point (t,t,t) will dramatically change the A-S prices: either all costs will be assigned to the first or the second output.

Moreover, note that A-S prices of zero result from the presence of *exterior facets* (added to the cost function through the additional assumption of free disposability) with partial derivatives of 0 as seen above.

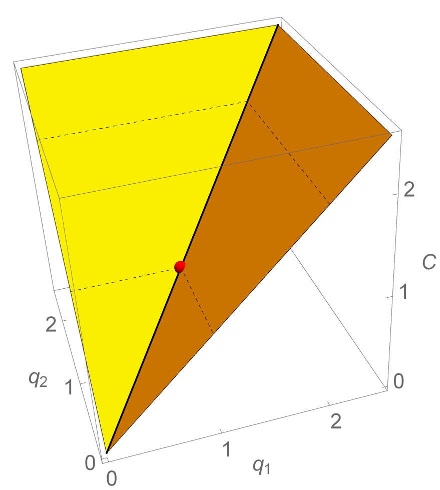

Figure 1: Simple case with multiple allocations in crs

Allowing for diseconomies of scale (as in drs) we encounter an additional problem. The cost estimate for points that are not weakly dominated by q^* is infinite. Intuitively, this follows from the fact that for points that are not dominated by our observation, we have no way to claim that production is feasible. The set of points with non-existing A-S prices in this case is composed of the shaded areas in Fig. 2. For such productions, small perturbations will not suffice to establish marginal costs.

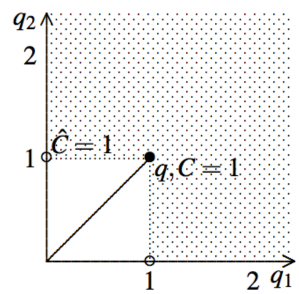

Figure 2: Simple case with multiple allocations in drs

△

Example 3 above identified three potential problems:

1. For some observations parts of the projected path from 0 to q may belong to intersecting hyperplanes causing multiple alternative solutions.

2. For some observations, some outputs may be associated with a zero A-S price.

3. For some data points A-S prices may be infinite relative to the estimated cost function $\widehat{C}$.

The first problem arises when a line segment $[aq, bq]$, $a < b,\ a, b \in [0, 1]$, belongs to multiple supporting hyperplanes simultaneously. One way to proceed in this case is therefore to induce an ordering of the products and to always allocate most costs to product 1, next most to product 2, etc. Such a principle serves to select which of the intersecting hyperplanes we shall use to derive the marginal costs. On such a hyperplane, marginal costs (and thereby A-S prices) are unique. Yet, if the chosen hyperplane is axis parallel it implies that one of A-S prices is zero (see also the illustrative example in Section 6). To select hyperplanes in case of multiple feasible hyperplanes, one can use lexicographic or pre–emptive goal programming. Ignizio (1976) gives an algorithm showing how a lexicographic goal program can be solved as a series of linear programs.[6]

[6]The idea is similar to the idea of using infinitesimal elements ε in a DEA program to eliminate slack, see e.g. Cooper, Seiford and Tone (2000).

To be more precise, consider the cost minimization problem associated with some output vector q^*:

$$
\begin{aligned}
\min \quad & \beta && (21)\\
\text{s.t.} \quad & \sum_{j=1}^{h} \lambda_j q_j \geq q^* && (22)\\
& \sum_{j=1}^{h} \lambda_j C_j \leq \beta && (23)\\
& \sum_{j=1}^{h} \lambda_j \leq 1 && (24)\\
& \lambda_j \geq 0, \forall j. && (25)
\end{aligned}
$$

where (24) is ignored if $\widehat{C}^{crs}$ is considered.

The dual program is to find inputs and output prices v and u ($u, v \geq 0$) such that

$$
\begin{aligned}
\max \quad & uq^* + \eta && (26)\\
\text{s.t.} \quad & v \leq 1 && (27)\\
& -vC^k + \sum_{i=1}^{n} u_i q_i^k + \eta \leq 0 \qquad \forall k = 1, \ldots, h && (28)\\
& \eta \leq 0 && (29)
\end{aligned}
$$

where (29) becomes an equality requirement in the constant returns to scale case.

We can reformulate this as a lexicographic goal programming problem by indicating the sequence in which the objectives shall be optimized:

$$
\begin{aligned}
\max \quad & P0(uq^* + \eta) + P1(u_1 q_1^*) + P2(u_2 q_2^*) + \cdots + Pn(u_n q_n^*) && (30)\\
\text{s.t.} \quad & v \leq 1 && (31)\\
& -vC^k + \sum_{i=1}^{n} u_i q_i^k + \eta \leq 0 \qquad \forall k = 1, \ldots, h && (32)\\
& \eta \leq 0 && (33)
\end{aligned}
$$

with the interpretation that we first optimize $P0$, then for fixed value of $P0$, we optimize $P1$, etc.

In practice, such an ordering of outputs may for instance be derived from various managerial priorities or pricing policies of the firm.

To ensure that all A-S prices are positive we may, for instance, use the extended facet approach of Olesen and Petersen (1996, 2003), but this will not be pursued further in the present paper.[7]

Finally, the third problem, i.e., that data points may not have finite cost estimates, is addressed by restricting attention to,

$$\mathscr{Q} = \{q^* \in \mathbf{R}^n_+ | \exists\, C^* : C^* \geq \widehat{C}(q^*)\} \tag{34}$$

i.e., to the set of feasible productions q given the observed data set $\mathscr{D}$. In the drs-case, this can be rewritten as

$$\mathscr{Q} = \{q^* \in \mathbf{R}^n_+ | \exists\, C^* : \sum_{j=1}^{h} \lambda_j q_j \geq q^*, \sum_{j=1}^{h} \lambda_j C_j \leq C^*, \sum_{j=1}^{h} \lambda_j \leq 1,\ \lambda_j \geq 0, \forall j\} \tag{35}$$

where $\sum_{j=1}^{h} \lambda_j \leq 1$ is ignored in the crs-case.

5 Inefficient Production

We will now consider how firms with inefficient production can be handled in the context of A-S cost allocation. Since A-S prices are only defined for efficient production we need specific assumptions about firm behavior in order to link the observed inefficiency to some underlying efficient counterpart. To do this transparently and in a consistent manner we will employ the notion of *rational inefficiency* introduced in Bogetoft and Hougaard (2003).

The main idea behind the notion of rational inefficiency is that what appears as inefficient production may actually benefit the firm. In particular, it is suggested that inefficiency may represent a form of indirect, on-the-job compensation to agents in the firm and therefore serves an important purpose as alternative means of payment. Hence, it is often reasonable to assume that inefficiency may be the result of a deliberate (rational) choice by the firm and not just a consequence of ignorance, bad organization of processes, wrong incentives etc. As such, we will generally think of firms as seeking not only to maximize profit by minimizing costs, but also to enjoy slack in the production process.[8]

[7] Using the extended facet approach would also solve the problem of infinite A-S prices.

[8] For further discussion about the various types and origins of inefficiency, see e.g., Bogetoft and Hougaard (2003).

Firms may introduce slack (inefficiency) on the input side as well as the output side. In the present cost context, the input side is represented by the aggregate (joint) cost and the introduction of slack on the input side naturally corresponds to assuming that firms deliberately allow for a constant fraction of overspending. The size of that fraction is revealed by the inefficiency level that the firm actually chose. This approach will be formalized in Subsection 5.1 below. In Subsection 5.2 we will subsequently consider the case where firms introduce slack on the output side. That is, for given costs, firms benefit from the lost (or rather self consumed) outputs. Specifically, we assume that firms actually produce at the revenue maximizing point (given output prices or preference weights set by a regulator) and when we observe something else we can directly infer the preferred slack mix by comparison with the allocatively efficient output mix. Thus, according to this assumption the relevent A-S prices are those associated with revenue maximizing production given (q^*, C^*) and estimate $\widehat{C}$. These A-S prices will generally differ from the preference weights set by the regulator and reflect the internal (unit) production costs of the firm.

Finally, note that our approach, and associated results, do not depend on the way we estimate the empirical cost function.

5.1 Cost inefficiency

The fundamentals are the observed output-cost data $(q^*, C^*) \in \mathscr{D}$ and the associated (estimated) cost function $\widehat{C}$. Note that $C^* \geq \widehat{C}(q^*)$ by definition of $\widehat{C}$. A strict inequality means that the firm is inefficient.

One way to think of inefficiency is simply as an excess use of inputs. Instead of the minimal cost $\widehat{C}(q^*)$, the firm has actually chosen to spend C^*. Following the notion of rational inefficiency we find it reasonable to assume that this tendency to use extra costs is not restricted to the particular output level q^*, i.e., that the firm would use a similar share of excess resources for any other production level as well. Indeed, if there would be efficiency gains from changing the size of production a rational firm would already have have done so. Note that efficient firms have revealed that they do not gain utility from excess use of resources.

Assumption 1: *(Rational Cost Inefficiency)* An inefficient firm with observed output-cost combination (q^*, C^*), in effect, faces a cost function

$$q \mapsto \frac{1}{E^c((q^*, C^*), \widehat{C})} \widehat{C}(q)$$

where E^c is the cost efficiency score given by (10).

Conceptually this means that we see the firm as deliberately spending a fraction $(1-E^c((q^*,C^*),\widehat{C}))$ in excess of what is necessary in order to produce any level of output q.

The idea that inefficiency shows up as a general scaling of the costs is convenient in the context of A-S pricing since in general, A-S prices are linear in the cost level, i.e.,

$$p^{AS}(q,\alpha C)=\alpha p^{AS}(q,C), \tag{36}$$

for $\alpha \in \mathbf{R}_+$.[9]

Assumption 1 allows us to use the Aumann-Shapley rule in case of inefficient production in a straightforward way.

Proposition 1: *Consider a set of observations $\mathscr{D}$ with specific inefficient observation (q^*,C^*). Given Assumption 1, we have for all outputs $i=1,\ldots,n$, that,*

$$\begin{aligned} \phi_i^{AS}(q^*,\tfrac{\widehat{C}}{E^c}) &= \tfrac{1}{E^c}\phi_i^{AS}(q^*,\widehat{C}) \\ &= \phi_i^{AS}(q^*,\widehat{C})+\tfrac{\phi_i^{AS}(q^*,\widehat{C})}{\sum_{j=1}^n \phi_j^{AS}(q^*,\widehat{C})}(1-E^c)C^*. \end{aligned}$$

Proof: The first equality follows from Assumption 1 and (36). Consider the second equality and a given output i. We have that

$$\frac{1}{E^c}\phi_i^{AS}(q^*,\widehat{C})=\frac{C^*}{\widehat{C}(q^*)}\phi_i^{AS}(q^*,\widehat{C})=(1+\frac{(1-E^c)C^*}{\widehat{C}(q^*)})\phi_i^{AS}(q^*,\widehat{C})=$$

$$\phi_i^{AS}(q^*,\widehat{C})+\frac{\phi_i^{AS}(q^*,\widehat{C})}{\sum_{j=1}^n \phi_j^{AS}(q^*,\widehat{C})}(1-E^c)C^*$$

since $\sum_{j=1}^n \phi_j^{AS}(q^*,\widehat{C})=\widehat{C}(q^*)$ by budget-balance. Q.E.D.

By Proposition 1 we see that focussing on cost efficiency (as in Assumption 1) is indeed natural since it opens up for an alternative way of looking at the problem: in fact, it gives the same result as if we were assuming that the level of cost inefficiency for output vector q^* can be construed as a general fixed cost $F=(1-E^c)C^*$ in production. As such, we can alternatively analyze the particular situation given by observation (q^*,C^*) as if a fixed cost $F=(1-E^c)C^*$ had been added to the cost function such that the efficient cost of any level q is given by $\widehat{C}(q)+F$.

[9] See e.g., Mirman and Tauman (1982).

Proposition 1 shows that scaling A-S cost shares of the cost efficient production $\widehat{C}(q^*)$ by a factor $1/E^c$ is tantamount to splitting the fixed cost of inefficiency F in proportion to these A-S cost shares and add this to the cost shares of cost efficient production.

The advantage of this result is furthermore that it works well with the way we handled fixed costs $C(0)$ in the case of vrs and irs technologies in Section 4. Here we also allocated fixed cost in proportion to the variable cost A-S shares. By straightforward combination of (20) and Proposition 1, we therefore obtain the following result.

Corollary 1: *Consider a set of observations $\mathscr{D}$ with a corresponding minimal extrapolation principle cost function $\widehat{C}^\gamma$, where γ is vrs, irs, drs or crs. Let*

$$\widetilde{C}^\gamma = \widehat{C}^\gamma - \widehat{C}^\gamma(0)$$

be the variable part of the cost function, and consider a specific inefficient observation (q^*, C^*). Given Assumption 1, we have for all outputs $i = 1, \ldots, n$, that,

$$\begin{aligned} \phi_i^{AS}(q^*, \tfrac{\widehat{C}^\gamma}{E^c}) &= \tfrac{1}{E^c}\phi_i^{AS}(q^*, \widehat{C}^\gamma) \\ &= \phi_i^{AS}(q^*, \widetilde{C}^\gamma) + \frac{\phi_i^{AS}(q^*, \widetilde{C}^\gamma)}{\sum_{j=1}^n \phi_j^{AS}(q^*, \widetilde{C}^\gamma)}[(1-E^c)C^* + \widehat{C}^\gamma(0)]. \end{aligned}$$

where E^c is the cost efficiency of (q^*, C^*) with respect to $\widetilde{C}^\gamma$.

In the case of crs, we know that cost efficiency E^c is the inverse of output efficiency E^q. We can therefore also make an alternative interpretation of the A-S cost shares with inefficiency. We record this as a Corollary.

Corollary 2: *Consider a set of observations $\mathscr{D}$ with specific inefficient observation (q^*, C^*). Given Assumption 1, in the crs-case we have that*

$$\begin{aligned} \phi^{AS}(q^*, \tfrac{\widehat{C}^{crs}}{E^c}) &= \tfrac{1}{E^c}\phi^{AS}(q^*, \widehat{C}^{crs}) \\ &= E^q\phi^{AS}(q^*, \widehat{C}^{crs}) \\ &= \phi^{AS}(E^q q^*, \widehat{C}^{crs}). \end{aligned}$$

Proof: By Assumption 1 and (36) we have

$$\phi^{AS}(q^*, \alpha\widehat{C}^{crs}) = p^{AS}(q^*, \alpha\widehat{C}^{crs})q^* = \alpha p^{AS}(q^*, \widehat{C}^{crs})q^* = \alpha\phi^{AS}(q^*, \widehat{C}^{crs}).$$

Thus, the first equality follows from the fact that $\widehat{C}^{crs}(q^*) = E^c C^*$. The second equality follows from (11). The third equality follows from the fact that $p^{AS}(q, C^{crs})$ is constant (and equal to the marginal cost): Hence, we have that $\phi^{AS}(E^q q^*, \widehat{C}^{crs}) = E^q q^* p^{AS}(E^q q^*, \widehat{C}^{crs}) = E^q q^* p^{AS}(q^*, \widehat{C}^{crs}) = E^q \phi^{AS}(q^*, \widehat{C}^{crs})$. Q.E.D.

In other words, the A-S cost shares of inefficient production can be found by multiplying the A-S cost shares of the corresponding efficient production with the output efficiency score E^q (given by (9)) or equivalently the inverse of the cost efficiency score E^c (given by (10)). Moreover, it does not matter whether we allocate the actual cost C^* using A-S prices related to q^* and scale up with a factor $1/E^c = E^q$ or the A-S-prices related to the efficiency adjusted output level $E^q q^*$ directly. Indeed, when the estimated cost function has constant returns to scale, the A-S-prices are the same along the entire ray through 0 and q^*.

5.2 Output Inefficiency

We now consider the case where firms introduce slack (inefficiency) on the output side. The main postulate here is that firms benefit from self consumed outputs, broadly interpreted. Let us assume that a firm with output-cost profile (q^*, C^*) is actually producing some other efficient output level $q \geq q^*$, but that only q^* is observed since the firm itself consumes $q - q^*$. In such cases, it would be natural to allocate costs based on the efficient production level q rather than the observed level q^*.

The problem with this line of thinking is of course that q is unobserved. Yet, by Proposition 1 and Corollary 1 in Bogetoft and Hougaard (2003), it is proved that a firm with preferences for both profit and slack will choose q as the allocatively efficient output mix.

Specifically, assume that output prices are given by the price vector p. This may be market prices or as in the empirical example provided in the next section, the prices (preference weights) set by a regulator. Given these prices, a rationally inefficient firm will choose output vector q as the allocatively efficient solution q^{AE} to

$$\max_q \quad p \cdot q \tag{37}$$

$$\text{s.t.} \quad \widehat{C}(q) \leq C^* \tag{38}$$

The intuition behind this result is that for a given cost level, the firms choice problem is really one of choosing the underlying production q so as to maximize revenue. This leaves the firm with maximal possibility to enjoy slack $q - q^*$.

Of course, if this vector difference is not non-negative in all dimensions, we can say that the firm has not behaved fully rational. Our assumption concerning rational output inefficiency can therefore be formulated as follows:

Assumption 2: (*Rational Output Inefficiency)* Given an output price vector p, an inefficient firm with observed output-cost combination $(q^*, C^*))$ chooses the underlying production plan q such as to maximize revenue subject to $q \geq q^*$, i.e., as the solution to,

$$\max_q \quad p \cdot q \tag{39}$$

$$\text{s.t.} \quad \widehat{C}(q) \leq C^* \tag{40}$$

$$q \geq q^* \tag{41}$$

Let $q^{AE}(p, \widehat{C}; q^*)$ solve the above programming problem.

Assumption 2 implies that when we observe output level q^*, there is an underlying production plan $q^{AE}(p, \widehat{C}; q^*)$, which the rational firm actually did produce (before consuming slack internally) and, which should therefore be used when determining the A-S prices and the associated A-S cost allocation.

To illustrate the idea, imagine a firm that produces q but internally consumes a large share of the first output q_1. It may for example be a baker that produces multiple types of bread but only consumes one of them himself. Now, if we allocate costs according to the observed production q^* (i.e., what is left of q after his own consumption), the cost of the bakers consumption will effectively be spread over all breads, but if we allocate the cost according to q, it will be allocated to the first type of bread. Another example may be a firm that have quality issues in the production process. Assume that a large share of the first product generally have to be discarded. If we allocate cost according to the set of products of good quality, all products will share the costs of the quality issue in the production of the first product. If instead we take the rational inefficiency approach, we will allocate the extra costs to the first product.

We now record the straightforward consequence of Assumption 2 with respect to A-S cost allocation.

Proposition 2: *Consider a set observations $\mathscr{D}$ with a corresponding minimal extrapolation principle cost function $\widehat{C}^\gamma$, where γ is vrs, irs, drs or crs, and consider a specific inefficient firm with observation (q^*, C^*). Given Assumption 2, we have that,*

$$\phi^{AS}(q^*, \widehat{C}^\gamma) = p^{AS}(q^{AE}(p, \widehat{C}^\gamma; q^*), \widehat{C}^\gamma) q^{AE}(p, \widehat{C}^\gamma; q^*).$$

for output price vector p.

Notice that with a non-parametric estimation of the cost function there may be cases with multiple allocatively efficient points for a given observation. The problem is of course that these points typically will have different A-S prices. In such cases we therefore suggest to apply the ordering of outputs similarly to the way we addressed the issue of multiple supporting hyperplanes for given efficient point in Section 4 above.

6 An Illustrative Example

Water companies are by and large natural monopolies that usually provide two services, namely the production of water and the distribution of water to households and firms. The water network requires large infrastructure investments making it economically optimal to only have one distribution network in each area. Water from different sources can in theory be distributed using the same network. Much like the electricity sector, it is therefore possible to have competition among producers of water and to reduce the natural monopoly scope to the distribution. Still, preservation of ground water reserves and quality concerns have in many jurisdictions led to natural geographical monopolies in water production as well.

In most countries, the regulation of water companies has been low powered, typically based on a cost plus regime. This seems to be changing. In recent years there have been several attempts to move towards a more high powered regime like a revenue cap similar to the regulation that has long been prominent for electricity distribution companies in for example Europe, cf., Bogetoft (2012).

When the typical operators are responsible for both production and distribution of water, it is interesting to allocate total costs among these activities. This may for example guide which tariffs different consumer types shall pay, and it may guide the access fees an incumbent waterworks with both production and distribution can charge external producers.

The first high powered regulation of Danish waterworks was introduced in 2012. The regulation is a benchmarking based revenue cap regulation. The benchmarking model is a crs DEA model with one main output, net volume. The net volume is the sum of several net volume elements that are designed to measure the amount of activities in different areas.

In the following we will use the 2011 data used to make the 2012 regulation and illustrate how this data can be used to allocate cost between water production and water distribution. There are 210 waterworks in the data set, and the amount of water production and water distribution activities (measured in netvolumes) as well as the operating costs (measured in

DKK) are summarized in Table 1 below.

	Mean	Std. dev	Min	Max
Water Production, q_1	2888260.25	8396426.53	0.00	109821313.10
Water Distribution, q_2	3916069.50	7607266.44	571.20	74436410.28
Cost, C	7026759.82	15449475.55	55073.21	184313172.69

Table 1: Descriptive statistics.

6.1 The Estimated Cost Function

We will focus on non-parametric estimation of the empirical cost functions in case of crs (as assumed by the Danish regulator) and drs for the sake of illustration. That is, the estimated cost function is given by the program (4)-(8) (in case of drs), excluding (7) in case of crs.

In case of crs we find only three efficient waterworks: Hjerting Vandværk Amba, Bjøvlund Vandværk, and, Nordenskov Vandværk. These three observations span the convex cone of the estimated cost function. Under drs there turns out to be additionally seven efficient waterworks spanning the cost function, all listed in Table 2 below.

Name	Water Production, q_1	Water Distribution, q_2	Cost, C
KE Vand A/S	109821313.10	74436410.28	184313172.69
Vandcenter Syd A/S	19986564.36	34731824.80	47964952.93
Sjælsø Vand A/S	15061136.50	109336.00	14367705.59
Hjerting Vandværk Amba	835293.97	3760657.20	1882561.43
Bjøvlund Vandværk	1179982.57	188126.60	737445.92
Nordenskov Vandværk	590749.90	593366.00	436994.91
Helle Vest Vandværk	1163503.15	1068388.20	869824.16
Vestforsyning Vand A/S	8307093.20	13291457.40	15500628.99
Århus Vand A/S	1163503.15	1068388.20	869824.16
Hjørring Vandselskab A/S	8586097.67	10675079.00	12526393.57

Table 2: Efficient waterworks (drs).

The estimated cost function assuming drs is illustrated graphically in Figure 3, where we have omitted the axis-parallel hyperplanes (arising from assumption A1).

6.2 A-S Cost Allocation with Rational Cost Inefficiency

In the following we shall allocate the total cost of each waterworks on the two outputs using the Aumann-Shapley allocation rule and our assumption of rational cost inefficiency

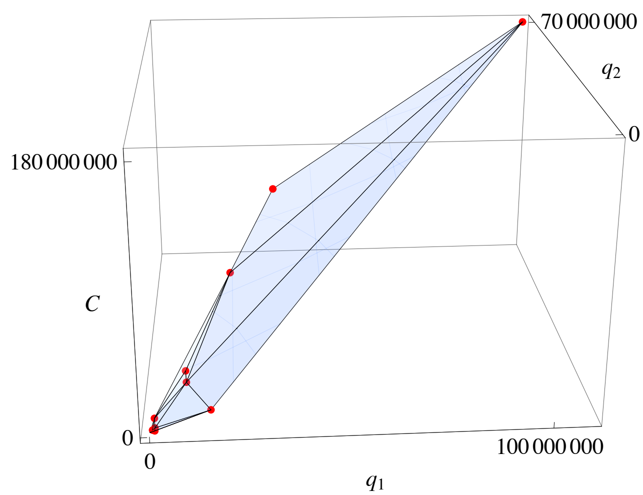

Figure 3: $\widehat{C}^{drs}$ based on the data set

(Assumption 1). In all cases the associated Aumann-Shapley prices are calculated using Matlab to solve the program (16)-(19) where t is gradually increased from 0 to 1. Since the constraints remain unchanged when going through the individual waterworks (only the objective function varies) we can improve the computational speed by first finding the set of vertices (extreme points of the frontier) defined by the intersecting hyperplanes given by the constraints. For each of the waterworks we then evaluate the objective function on all those vertices (since the solution will always lie in a vertex). In this way we obtain two advantages; first, we can easily detect how many hyperplanes the production plan in question lies on, and second, we get solutions significantly faster than by solving the LP problem for each of the waterworks.

Turning to the drs-case: Obviously no waterworks will face the problem of infinite A-S prices since they are all included in the estimation of the cost function. Three of the efficient waterworks face the problem of multiple alternative solutions. For instance this is the case for Nordenskov Vandværk, which is illustrated in Figure 4.

The green line lies on two hyperplanes simultaneously, and the two possible candidates for A-S prices are $p^{AS} = (0.6033, 0.1338)$ and $p^{AS} = (0.3050, 0.4329)$, respectively. In both cases all costs are allocated to the both outputs, because Nordenskov Vandværk is efficient. To choose between the two possible A-S price candidates one can define an exogenous ordering of outputs as described in section 4.

If most costs are allocated to output 1 and the A-S allocation becomes:

$x_1^{AS} = 356405.15$

$x_2^{AS} = 80589.76$

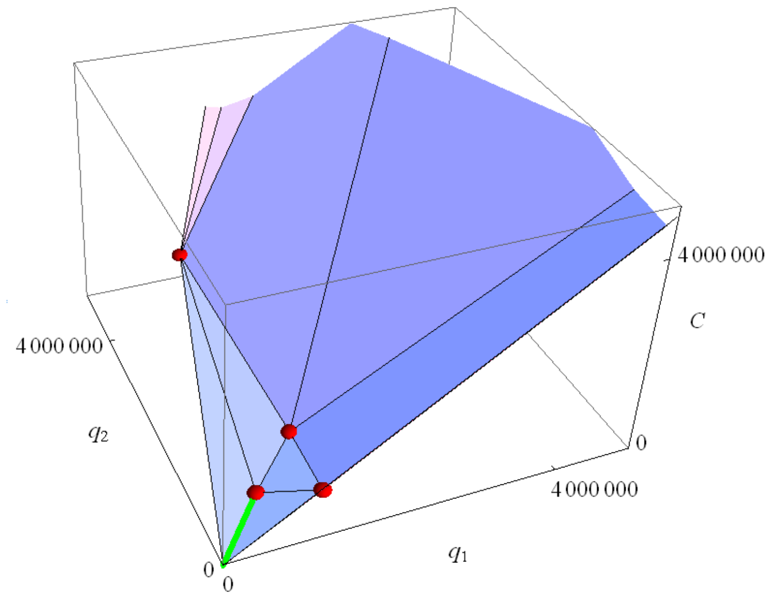

Figure 4: Cost Possibility Set (zoomed in)

If most costs are allocated to output 2 we get:
$x_1^{AS} = 180150.79$
$x_2^{AS} = 256844.12$

Another efficient waterworks that faces the same problem is Hjerting Vandværk Amba. The two possible candidates for A-S prices are $p^{AS} = (0.3050, 0.4329)$ or $p^{AS} = (0, 0.5006)$ since now one of the supporting hyperplanes is axis parallel (creating an exterior facet). So one can choose between a case where both A-S prices are strictly positive and one where one price is zero. Here, it seems most reasonable to avoid the solution with a zero A-S price. Figure 5 illustrates.

From the entire sample of 210 waterworks, 21 face the problem of having A-S prices of zero if a specific ordering of the outputs is specified. Apart from Hjerting Vandværk Amba mentioned above, Bjøvlund Vandværk is the only one where there is an alternative strictly positive A-S price vector.

As mentioned, zero-prices appears when a supporting hyperplane is axis-parallel. This is illustrated in figure 6. The brown points are all observations where one of the outputs has A-S price zero.

In Figure 7 the costs of the inefficient waterworks are scaled down to efficient costs.

Next, we focus on A-S allocation for inefficient waterworks. Consider the case of Arwos Vand A/S.

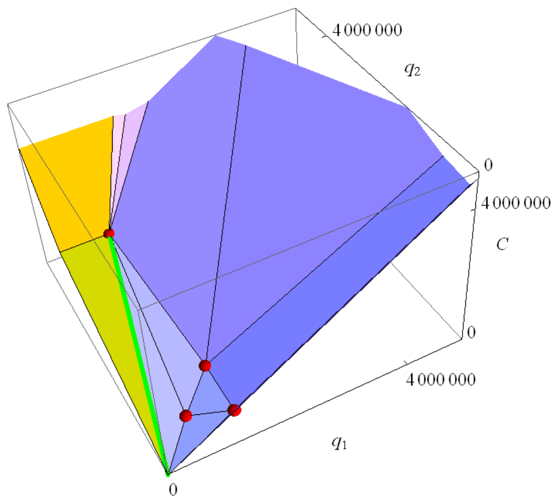

Figure 5: Cost Possibility Set (zoomed in)

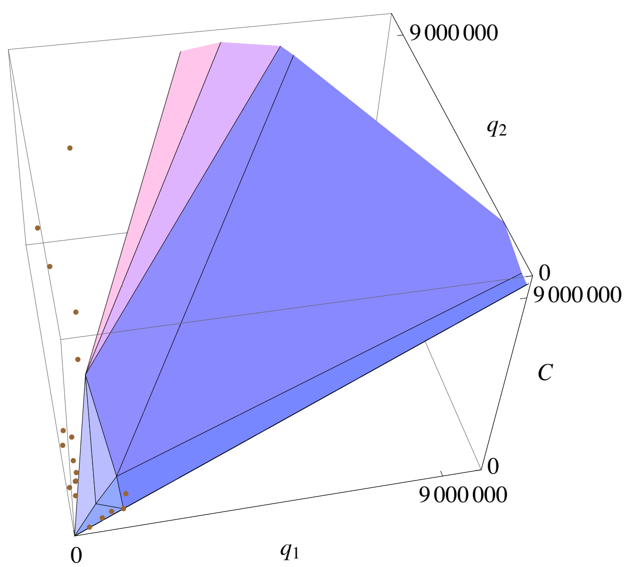

Figure 6: Cost Possibility Set

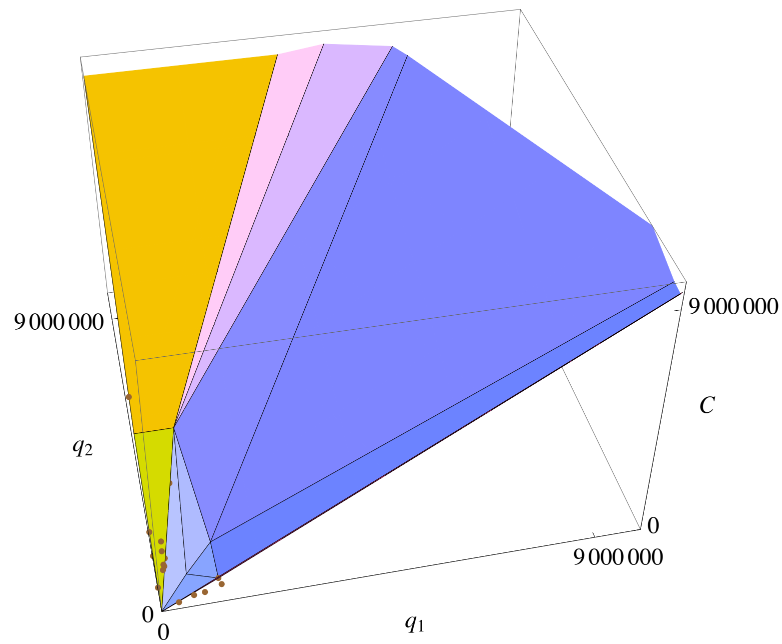

Figure 7: Cost Possibility Set

Name	Water Production	Water Distribution	Cost
Arwos Vand A/S	2869541.62	3541004.80	8218777.91

First we find the A-S prices in the (hypothetical) situation where the production is cost efficient (i.e., with a total cost of 3679269.17). These prices are $p^{AS} = (0.7066, 0.4664)$ and the A-S allocation of (efficient) cost hence becomes;

$x_1^{AS} = 2027728.22$
$x_2^{AS} = 1651540.96$

The remaining costs are inefficiency costs, which are allocated in proportion to the A-S cost shares (cf., Proposition 1). For instance output 1 gets allocated an additional inefficiency cost of

$$\frac{\phi_1^{AS}}{\phi_1^{AS}+\phi_2^{AS}}(C^* - C^* E^c) = \frac{2027728.22}{2027728.22+1651540.96}(8218777.91 - 3679269.17)$$

$$= 2734589.54.$$

The final allocation thus becomes:

$x_1^{AS} = 4529553.86$
$x_2^{AS} = 3689224.05$

In the drs-case there are 200 inefficient waterworks with an average cost inefficiency of

2262943.92 corresponding to 50.95 % of the total costs. No inefficient waterworks face the problem of multiple A-S prices, and on average 40.36 % of the total costs are allocated on output 1 and 59.64 % on output 2.

6.3 A-S Cost Allocation with Rational Output Inefficiency

Turning to output inefficiency, we shall assume that the relevant output prices are the same for both outputs. The regulator effectively construct the total net volume by adding the net volume of production and distribution. This means that the waterworks will use 1:1 prices on the outputs when they try to “play the regulation”. Using 1:1 prices, we can calculate the productions that maximize the revenue permitted by the regulator.

According to Assumption 2, the revenue maximizing allocatively efficient production plan is the underlying production plan chosen by the waterworks. The difference between this production plan and the observed production plan is the slack ”consumed” by the waterworks.

In the case of Arwos we find the allocatively optimal production as

$$(q_1, q_2, C) = (5449306.06, 7876775.60, 8218777.91),$$

solving (37).

The corresponding A-S prices are $p^{AS} = (0.8173, 0.4780))$ and hence the resulting A-S cost allocation becomes:

$x_1^{AS} = 4453671.04$

$x_2^{AS} = 3765120.09$

which is rather close to the allocation found by considering cost inefficiency (as is indeed the case for the ”average observation”, cf., table below). The graphical interpretation of input inefficiency and rational output inefficiency is illustrated in figure 8 for Arwos.

In other cases the results of using Rational Input and Output Inefficiency can differ much more as shown in the following tables where absolute as well as relative differences in allocated cost shares are found between cost and output inefficiency. If outputs are ordered $1 \succ 2$ we get:

	Output 1 (abs)	Output 2 (abs)	Output 1 (rel)	Output 2 (rel)
mean	-185009.53	185411.74	0.20%	-0.20%
min	-6976356.43	-8267539.27	-81.71%	-71.60%
max	8268821.72	6977146.66	71.60%	81.71%

If outputs are ordered $2 \succ 1$ we get:

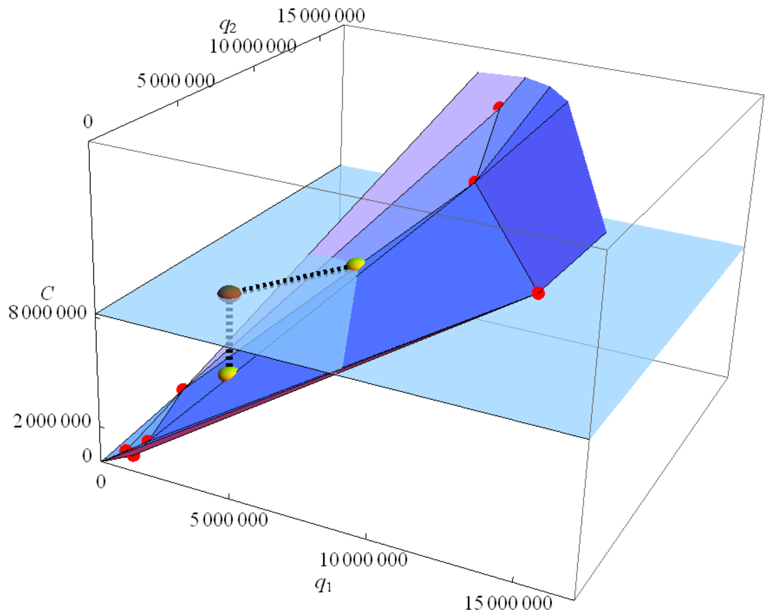

Figure 8: Arwos, inefficiency

	Output 1 (abs)	Output 2 (abs)	Output 1 (rel)	Output 2 (rel)
mean	-182594.78	182996.98	1.01%	-1.01%
min	-6976356.43	-8267539.27	-81.71%	-71.60%
max	8268821.72	6977146.66	71.60%	81.71%

Consider, for instance, the waterworks Mariager Vand Amba, which has the output cost profile $(q_1, q_2, C) = (150599.20, 811534.80, 642263.95)$. When using input cost inefficiency we see that the same output could be obtained at cost 406249.24, and the final allocation of actual cost (for both orderings of outputs) is:

$x_1^{AS} = 0$

$x_2^{AS} = 642263.95$

This highlight the problem of allocating zero costs. If instead we use the rational output inefficiency approach with price ratio 1:1, we get that the following allocatively efficient point: $(q_1, q_2, C) = (862377.82, 818645.02, 642263.95)$. The final allocation (for both orderings of outputs) therefore is:

$x_1^{AS} = 524813.62$

$x_2^{AS} = 117443.18$

The absolute difference here is $0 - 524813.62 = -524813.62$ for output 1 and $642263.95 - 117443.18 = 524820.77$ for output 2 (the slight difference is caused by rounding errors), corresponding to relative differences of -81.71% and 81.71% respectively. When using rational output inefficiency no waterworks face the problem of allocating zero costs.

It is not surprising, as such, that the difference between A-S cost shares with respect to cost and output efficiency can be rather big. Clearly, when deviating from an overall assumption of constant returns to scale, the efficiency of certain observations can be very different in cost and output space respectively. In a specific application, like the one above, the choice between cost (input) or output orientation is in many ways a counterpart of the similar type of choice in a conventional efficiency analysis: if focus is on cost savings and inefficiency is mainly due to ”bad” utilization of resources it seems natural to allocate costs using A-S prices associated with cost minimization; if focus is on quality issues or other forms of ”slack” allocation in production it seems natural to employ A-S prices associated with output efficient production. As such the decision is ad hoc and related to the data set at hand.

7 Final Remarks

In the efficiency measurement literature it is well recognized that the non-parametric estimation of the efficient frontier (of either the production or cost function) is sensitive to small changes in the data since the frontier is spanned by ”extreme” observations. Obviously non-parametric estimation of A-S prices inherits this sensitivity as the gradients on the projected path (from 0 to q) may change dramatically moving from one efficient facet of the convex polyhedral to another. The remedy is usually to bootstrap the estimated function, see e.g. Simar and Wilson (1998). The sensitivity of A-S price estimates and the use of bootstrapping techniques is a topic we leave for future research

References

[1] Asmild, M, P. Bogetoft and J.L. Hougaard (2009), Rationalising DEA estimated inefficiencies, *Journal of Business Economics ZfB*, 87- 97.

[2] Asmild, M, P. Bogetoft and J.L. Hougaard (2013), Rationalising inefficiency: Staff utilisation in branches of a large Canadian bank, *Omega*, 41, 80-87.

[3] Aumann, R and L. Shapley (1974), Values of Non-atomic Games, Princeton University Press.

[4] Banker, R.D. (1999), Studies in Cost Allocation and Efficiency Evaluation, UMI-Press.

[5] Banker R.D., A. Charnes and W.W. Cooper (1984), Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis, *Management Science*, 30, 1078-1092.

[6] Bazaraa, M.S., J.J. Jarvis and H.D. Sherali (1990), Linear Programming and Network Flows (2'ed.), Wiley.

[7] Billera, L.J. and C. Heath (1982), allocation of shared costs: a set of axioms yielding a unique procedure, *Mathematics of Operations Research*, 7, 32-39.

[8] Billera, L.J., C. Heath and J. Raanan (1978), Internal telephone billing rates - a novel application of non-atomic game theory, *Operations Research*, 26, 956-965.

[9] Bjørndal, M. and K. Jörnsten (2005), An Aumann-Shapley approach to cost allocation and pricing in a supply chain; In Koster and Delfman (eds), Supply Chain Management, CBS-Press, p. 182-198.

[10] Bogetoft, P. (1996), DEA on relaxed convexity assumptions, *Management Science*, 42, 457-465.

[11] Bogetoft, P. (1997), DEA-Based Yardstick Competition: The Optimality of Best Practice Regulation, *Annals of Operations Research*, 73, 277-298.

[12] Bogetoft, P., and J.L. Hougaard (2003), Rational inefficiency, *Journal of Productivity Analysis*, 20, 243-271.

[13] Bogetoft, P. and L. Otto (2010), Benchmarking with DEA, SFA and R, Springer New York.

[14] Bogetoft, P. J.M. Tama and J. Tind (2000), Convex input and output projections of nonconvex production possibility sets, *Management Science*, 46, 858-869.

[15] Bogetoft, P. (2012), Performance Benchmarking: Measuring and Managing Performance, Springer New York.

[16] Castano-Pardo, A. and A. Garcia-Diaz (1995), Highway cost allocation: An application of the theory of non-atomic games, *Transportation Research*, 29, 187-203.

[17] Charnes, A. W.W. Cooper and E. Rhodes (1978), Measuring the efficiency of decision making units, *European Journal of Operational Research*, 2, 429-444.

[18] Cooper, W.W., L.M. Seiford and K. Tone (2000), Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software, Kluwer Academic Publishers, Boston.

[19] Delis, M.D., M. Iosifidi and E. Tsionas (2014), On the estimation of marginal cost, *Operations Research*, 62, 543-556.

[20] Friedman, E. and H. Moulin (1999), Three methods to share joint costs or surplus, *Journal of Economic Theory*, 87, 275-312.

[21] Haviv, M. (2001), The Aumann-Shapley price mechnism for allocating congestion costs, *Operations Research Letters*, 29, 211-215.

[22] Hougaard, J.L. (2009), An Introduction to Allocation Rules, Springer.

[23] Hougaard, J.L. and J. Tind (2009), Cost allocation and convex data envelopment, *European Journal of Operational Research*, 194, 939-947.

[24] Ignizio, J.P. (1976), Goal Programming and Extensions, Lexington Books, Lexington, MA.

[25] Mirman, L. and Y. Tauman (1982), Demand compatible equitable cost sharing prices, *Mathematics of Operations Research*, 7, 40-56.

[26] Mirman, L., Y. Tauman and I. Zang (1985a), On the use of game-theoretic concepts in cost accounting; Chapter 3 in Young (Editor) Cost Allocation: Methods, Principle, Applications, North Holland.

[27] Mirman, L., Y. Tauman and I. Zang (1985b), Supportability, sustainability and subsidy-free prices, *RAND Journal of Economics*, 16, 114-126.

[28] Mirman, L., D. Samet and Y. Taumann (1983), An axiomatic approach to allocation of a fixed cost through prices, *Bell Journal of Economics*, 14, 139-151.

[29] Olesen, O.B. and N.C. Petersen (1996), Indicators of ill-conditioned data sets and model specification in Data Envelopment Analysis: an extended facet approach, *Management Science*, 42, 205-219.

[30] Olesen, O.B. and N.C. Petersen (2003), Identification and use of efficient faces and facets in DEA, *Journal of Productivity Analysis*, 20, 323-360.

[31] Samet, D., Y. Tauman and I. Zang (1984), An application of the Aumann-Shapley prices for cost allocation in transportation problems, *Mathematics of Operations Research*, 9, 25-42.

[32] Spulber, D.F. (1989), Regulation and Markets, MIT-Press.

[33] Simar, L. and P. Wilson (1998), Sensitivity analysis of efficiency scores: How to bootstrap in non-parametric frontier models, *Management Science*, 44, 49-61.

[34] Tsanakas, A. and C. Barnett (2003), Risk capital allocation and cooperative pricing of insurance liabilities,*Insurance: Mathematics and Economics*, 33, 239-254.

[35] Young, H.P. (1985), Producer incentives in cost allocation, *Econometrica*, 53, 757-765.