October 2013

WORKING PAPER SERIES

2013-ECO-22

Nonparametric cost and revenue functions under constant economies of scale: An enumeration approach for the single output or input case

Walter Briec
LAMPS, Université de Perpignan

Kristiaan Kerstens
CNRS-LEM and IESEG School of Management

Ignace Van de Woestyne
Hogeschool-Universiteit Brussel

IESEG School of Management
Lille Catholic University
3, rue de la Digue
F-59000 Lille
www.ieseg.fr
Tel: 33(0)3 20 54 58 92
Fax: 33(0)3 20 57 48 55

Nonparametric cost and revenue functions under constant economies of scale: An enumeration approach for the single output or input case[†]

Walter Briec[*]

Kristiaan Kerstens[**]

Ignace Van de Woestyne[***]

Abstract:

This note shows how the linear programs needed to compute cost and revenue functions under constant returns to scale and a single output or input, respectively, can be replaced with a more efficient enumeration algorithm. A numerical example illustrates this algorithm.

Keywords: nonparametric cost and revenue functions, enumeration, linear programming.

* LAMPS, Université de Perpignan, France.

** CNRS-LEM (UMR 8179), IESEG School of Management, 3 rue de la Digue, F-59000 Lille, France. k.kerstens@ieseg.fr Corresponding author.

*** Hogeschool-Universiteit Brussel, Brussels, Belgium.

July 2013

[†] This is a revised version of LEM - Document de travail 2011-11.

1. INTRODUCTION

Nowadays, cost and revenue functions are often estimated using nonparametric, deterministic estimators (see, e.g., Cooper, Seiford and Tone (2006), Hackman (2008), or Ray (2004)). This involves the computation of one linear program (LP) per observation under evaluation in the sample. Obtaining statistical inference from these extremum estimators using recent bootstrapping techniques requires again solving a LP in each draw (see, e.g., De Borger, Kerstens and Staat (2008) for an application). This can result in a substantial computational burden.

It has gone unnoticed so far that the computation of the cost function can be simplified in the single output case for constant economies of scale. Similarly, the solution of the revenue function also simplifies in the single input case under identical economies of scale. To the best of our knowledge this is the first contribution showing that an enumeration algorithm works for these specific convex Data Envelopment Analysis (DEA) value-based models.

Our contribution must be seen against the background of a small, burgeoning literature focusing on a variety of strategies to speed up the LP computations underlying DEA production frontier models. Ali (1993) is probably the first study initiating this research into the computational aspects of DEA. Following a taxonomy introduced in some early overview articles of Dulá (2002), one can distinguish between preprocessors, enhanced procedures, and new algorithms. In contrast to this rather substantial literature, to our knowledge very few articles have focused on simplifying the computational burden for computing cost or revenue functions. Following up on an earlier contribution by Camanho and Dyson (2005), Jahanshahloo, Soleimani-damaneh and Mostafaee (2008) simplify the LP formulations for traditional convex cost functions by cutting down on the amount of constraints and decision variables. Paryab, Khanjani Shiraz, and Jalalzadeh (2012) also manage to reduce the complexity of the LP formulations of nonconvex cost functions with various returns to scale assumptions by similarly reducing both the number of constraints and decision variables. Our approach continues this line of research by focusing on a specific returns to scale assumption on a convex technology and by restricting the numbers of inputs or outputs.

The purpose of this note is to prove both results regarding the use of enumeration for cost and revenue functions under constant returns to scale and a single output or input, respectively. Furthermore, apart from a general argument as to the computational complexity of enumeration versus LP, this note provides a small empirical illustration. Section 2 introduces basic definitions. Section 3 contains the main results. Section 4 illustrates using a numerical example. A concluding section offers some further perspectives.

2. TECHNOLOGY, COST AND REVENUE FUNCTIONS

Deterministic, nonparametric technologies are based on activity analysis. A technology uses a vector of inputs $x \in \mathbb{R}_+^N$ to produce a vector of outputs $y \in \mathbb{R}_+^M$. This technology or production possibility set is the set of all feasible input-output vectors: $T = \{(x,y): x \text{ can produce } y\}$. Alternatively, the input set $L(y)$ denotes all input vectors x producing the output vector y: $L(y) = \{x: (x,y) \in T\}$. Equally so, the output set $P(x)$ is defined as the set of all output vectors y that can be obtained from the input vector x: $P(x) = \{y: (x,y) \in T\}$.

The standard radial input efficiency measure is defined as:

$$DF_i(x,y) = \min \{\lambda \mid \lambda \geq 0, (\lambda\, x) \in L(y)\}. \qquad (1)$$

Its main properties are: (i) $0 < DF_i(x,y) \leq 1$, with efficient production on the boundary (isoquant) of $L(y)$ represented by unity; (ii) it has a cost interpretation (see, for instance, Hackman (2008)).

Assume that p is a vector of strictly positive input prices ($p \in \mathbb{R}_+^N$). Then, the cost function corresponding to a given technology is defined as follows:[1]

$$C(p,y) \quad = \quad \inf \{p \cdot x : \quad x \in L(y)\}. \qquad (2)$$

Figure 1 shows an isoquant of an input set $L(y)$ with two inputs generating the same level of outputs. The figure also contains an iso-cost line tangent to this isoquant. For a given observation, the radial distance to the isoquant represents its technical efficiency. The radial distance to the iso-cost line represents a measure of cost efficiency. Finally, since cost efficiency is always lower or equal to technical efficiency, in case there is a difference this can be attributed to allocative efficiency. The resulting basic efficiency decomposition states that cost efficiency is the product of a technical efficiency component and an allocative efficiency component (see Cooper, Seiford and Tone (2006, ch. 8) for further details).

< FIGURE 1 ABOUT HERE >

Equally so, assume that r is a vector of strictly positive output prices ($r \in \mathbb{R}_+^M$), then the revenue function corresponding to a given technology is defined by:

$$R(r,x) \quad = \quad \max \{r \cdot y : \quad y \in P(x)\}. \qquad (3)$$

Apart from imposing traditional assumptions on technology (i.e., no free lunch and inaction, closedness, free disposal of inputs and outputs, and convexity), the sole key assumption we invoke in this contribution is constant returns to scale (CRS) (i.e., *when* $(x,y)\in T$, *then* $\delta(x,y)\in T$, $\forall\delta > 0$). Several nonparametric technologies have been derived from these axioms (Banker, Charnes and Cooper (1984) are among the earlier sources).

A convex technology based on K observations (x_k, y_k), $k=1,...,K$, satisfying the above axioms and CRS has been defined in Charnes, Cooper and Rhodes (1978) as follows:

$$T^{\text{CRS}} = \{(x,y) \mid x \in \mathbb{R}_+^N,\ y \in \mathbb{R}_+^M,\ \sum_{k=1}^{K} z_k y_k \geq y\ ,\ \sum_{k=1}^{K} z_k x_k \leq x\ ,\ z_k \geq 0,\ k=1,...,K\}. \quad (4)$$

This article introducing this technology is considered to mark the start of the DEA literature. Computing a cost (1) or revenue (2) function with respect to this CRS technology is a standard model in the DEA literature (e.g., Cooper, Seiford and Tone (2006) or Cooper, Seiford and Zhu (2011)) and normally requires solving one LP per observation (eventually a simplified version as elaborated by Jahanshahloo, Soleimani-damaneh and Mostafaee (2008)).

3. MAIN RESULTS

Minimal assumptions on observed inputs and outputs are usually formulated as follows. Summing over all observations, there is a strictly positive aggregate production of every output and a strictly positive aggregate consumption of every input. Every unit produces a positive amount of at least one output and employs a positive amount of at least one input (see, e.g., pages 44-45 in Färe, Grosskopf and Lovell (1994)). When considering a single output case, this implies that all observations have a strictly positive single output. Likewise, for the single input case, this implies that all observations use a strictly positive single input.

Proposition 1: *In the case of CRS and a single, strictly positive output (M=1), the cost function $C^{CRS}(p,y)$ is computed as follows:*

$$C^{CRS}(p,y) = y \min_{k=1...K} \left\{ \frac{1}{y_k} \cdot p \cdot x_k \right\}.$$

Proof: Assume there is a single, strictly positive output (M=1). Consider the technology T^{CRS} enveloping the sample $S = \{(x_1, y_1),...,(x_K, y_K)\}$. For k = 1,…,K, denote $\delta_k = \frac{y}{y_k}$. Now,

[1] The radial input efficiency measure, being the inverse of the input distance function, is related to the cost

define the transformed sample $S' = \{(\delta_1 x_1, \delta_1 y_1), \ldots, (\delta_K x_K, \delta_K y_K)\} = \{(\delta_1 x_1, y), \ldots, (\delta_K x_K, y)\}$ realizing the same technology T^{CRS}. Denote the following sets:

$$L^{CRS}(y) = \left\{ x \in \mathbb{R}_+^N : \quad x \geq \sum_{k=1}^{K} z_k \delta_k x_k, \sum_{k=1}^{K} z_k \geq 1, z_k \geq 0 \right\};$$

$$L^{VRS}(y) = \left\{ x \in \mathbb{R}_+^N : \quad x \geq \sum_{k=1}^{K} z_k \delta_k x_k, \sum_{k=1}^{K} z_k = 1, z_k \geq 0 \right\}.$$

First, we demonstrate that $L^{CRS}(y) = L^{VRS}(y)$. Clearly, $L^{VRS}(y) \subseteq L^{CRS}(y)$. Conversely, let $x \in L^{CRS}(y)$. Then $x \geq \sum_{k=1}^{K} z_k \delta_k x_k$ for arbitrary $z_k \geq 0$ ($k = 1,\ldots,K$) with $\alpha = \sum_{k=1}^{K} z_k \geq 1$. Now, let $z_k' = \frac{z_k}{\alpha}$. Then, $x \geq \alpha \sum_{k=1}^{K} z_k' \delta_k x_k \geq \sum_{k=1}^{K} z_k' \delta_k x_k$ with $\sum_{k=1}^{K} z_k' = 1$. Thus, $x \in L^{VRS}(y)$ leading to the desired result. Second, $L^{CRS}(y) = L^{VRS}(y) = Co(\{\delta_1 x_1, \ldots, \delta_K x_K\}) + \mathbb{R}_+^N$, where $Co(.)$ denotes the convex hull. Since $Co(\{\delta_1 x_1, \ldots, \delta_K x_K\})$ is a convex polyhedron by definition, the minimum of any non-decreasing linear function (in casu, the cost function) is achieved at some vertex point (see Eremin (2002)). Thus,

$$C^{CRS}(p, y) = \min\{p \cdot x : \quad x \in L^{CRS}(y)\} = \min_k \{p \cdot \delta_k \cdot x_k\} = \min_k \left\{ \frac{y}{y_k} \cdot p \cdot x_k \right\}.$$

Q.E.D.

Remark: Using the formulation of the cost function in Camanho and Dyson (2005) (their formula (3)) with the conditions of Proposition 1, it can be easily shown that the optimal solution is obtained by using exactly one observation. Therefore, this optimal solution can be found by minimizing inner products with the price vector over all observations, thereby providing an alternative proof of Proposition 1.[2] However, this proof strategy does not work for the revenue function.

Proposition 2: *In the case of CRS and a single, strictly positive input (N=1), the revenue function* $R^{CRS}(r, x)$ *is computed as follows:*

$$R^{CRS}(r, x) = x \max_{k=1\ldots K} \left\{ \frac{1}{x_k} \cdot r \cdot y_k \right\}.$$

function via duality relation (see Hackman (2008) for details).

[2] We are grateful to a referee to explicitly outline this alternative way of proof.

Proof: Assume there is a single, strictly positive input (N=1). Consider a technology T^{CRS} enveloping the sample $S=\{(x_1,y_1),...,(x_K,y_K)\}$. For k = 1,…,K, denote $\mu_k=\frac{x}{x_k}$. Now, define the transformed sample $S'=\{(\mu_1 x_1,\mu_1 y_1),...,(\mu_K x_K,\mu_K y_K)\}=\{(x,\mu_1 y_1),...,(x,\mu_K y_K)\}$ realizing the same technology T^{CRS}. Using similar arguments as in Proposition 1, we obtain $P^{CRS}(x)=\left(Co\left(\{\mu_1 y_1,...,\mu_K y_K\}\right)+(-\mathbb{R}^M_+)\right)\cap\mathbb{R}^M_+=\left\{y\in\mathbb{R}^M:0\le y\le\sum_{k=1}^K z_k\mu_k y_k,\sum_{k=1}^K z_k=1,z_k\ge 0\right\}.$

Since the price vector $r\in\mathbb{R}^M_{++}$ is nonnegative, we have:

$$\begin{aligned}R^{CRS}(r,x)&=\max\{r\cdot y:y\in P^{CRS}(x)\}\\&=\max\left\{r\cdot y:y\in\left(Co\left(\{\mu_1 y_1,...,\mu_K y_K\}\right)+(-\mathbb{R}^M_+)\right)\cap\mathbb{R}^M_+\right\}\\&=\max\left\{r\cdot y:y\in\left(Co\left(\{\mu_1 y_1,...,\mu_K y_K\}\right)+(-\mathbb{R}^M_+)\right)\right\}\\&=\max\left\{r\cdot y:y\in Co\left(\{\mu_1 y_1,...,\mu_K y_K\}\right)\right\}.\end{aligned}$$

Since $Co\{(\mu_1 y_1,...,\mu_K y_K)\}$ is a convex polyhedron by definition, the maximum of any non-decreasing linear function (in casu, the revenue function) is achieved at some extreme point. Thus,

$$\begin{aligned}R^{CRS}(r,x)=\max\{r\cdot y:y\in P^{CRS}(x)\}&=\max\left\{r\cdot y:y\in Co\left(\{\mu_1 y_1,...,\mu_K y_K\}\right)\right\}\\&=\max_k\{r\cdot\mu_k\cdot y_k\}=\max_k\left\{\frac{x}{x_k}\cdot r\cdot y_k\right\}.\end{aligned}$$

Q.E.D.

We include an algorithm for computing C^{CRS} for all observations:

Algorithm 1:

For $i=1...K$ do:

1) Select the i^{th} observation $(x_i,y_i)=(x_{i1},...,x_{iN},y_{i1})$ and its input price vector $p_i=(p_{i1},...,p_{iN})$.
2) Put $C=\infty$.
3) For $k=1...K$ do:
   a. $C1=\frac{y_{i1}}{y_{k1}}\cdot\sum_{j=1}^N p_{ij}x_{kj}$
   b. If $C1<C$ then $C=C1$
4) The variable C holds the value of $C^{CRS}(p_i,y_i)$ for the i^{th} observation.

A similar algorithm could be formulated for R^{CRS} .

Having proven the two main results, we spell out the computational consequences in the next corollary.

Corollary 1: *In case of CRS and a single output, the cost function can be computed by enumeration in a smaller number of operations compared to LP. The same applies to the revenue function in case of CRS and a single input.*

Proof: In case of a single output, enumeration requires $O(LK(1+N)^2)$ arithmetic operations, where L is a measure of data storage for a given precision. Ignoring the worst case exponential complexity of the simplex method in LP, the Kamarkar interior point (IP) method needs $O(L(n)^{3.5})$ operations (with n the number of decision variables) while the most successful IP method known so far (i.e., primal-dual Newton step IP method) has a complexity of $O(L(n)^3)$ (see Chong and Zak (2001) or Eiselt and Sandblom (2007) for details). Transposed to our models, one thus needs at best $O(L(K+N)^3)$ operations for LP. Since in general $K > N \geq 1$, it follows that $K + N > 1 + N$ and consequently $(K + N)^2 > (1 + N)^2$. Also $K + N > K$ which combined with the previous inequality leads to $(K + N)^3 > K(1 + N)^2$. Hence, enumeration of the cost function under CRS and a single output is always quicker compared to LP. The same argument applies to Proposition 2.

Q.E.D.

While the above corollary is rather obvious from a computational point of view, it is good to put this basic result in context. First, there are the often cited rules of thumb in the frontier literature stressing that certain relations between the number of observations and the number of variables should be observed. For instance, Vassiloglou and Giokas (1990, p. 593) suggest that the sample should have at least twice as many observations as there are model variables. Second, recent insights into the statistical properties of frontier estimators show that these are consistent (with a slow rate of convergence because of the curse of dimensionality), but inherently biased towards unity (see, e.g., Fried, Lovell and Schmidt (2008)). This bias depends on specific properties of the underlying data: (i) number of observations in the sample, (ii) the number of inputs and outputs in the model, and (iii) the density of observations around the relevant segment of the frontier. Hence, for all these reasons practitioners ideally must seek to have sample sizes much larger than the number of inputs (K>>N), resulting in substantial gains in computer time for the newly proposed method.

The numerical and empirical illustration to which we now turn serves to document how substantial these gains may well turn out to be in practice.

4. NUMERICAL ILLUSTRATION

To illustrate the use of Proposition 1, we provide an artificial example containing six observations, each having two inputs x_1 and x_2 and one output y. Data is provided in the first four columns of Table 1. We assume a fixed unit price of 4 for x_1 and 5 for x_2.

Table 1: Cost Function Computations for Observation 1 with Fixed Input Price Vector

Obs	x_1	x_2	y	$C1$
1	2	2	2	18.00
2	3	2.5	4	12.25
3	2	4	5	$\boxed{11.20}$
4	0.5	4.5	3.5	14.00
5	5	5	7	12.86
6	4	4.5	4	19.25
Unit price	4	5		

The given observations and their resulting CRS efficient frontier are also visualized in three dimensions in Figure 2. Notice that the three CRS efficient observations (visible as clear red circles) are located on the frontier, while the other inefficient observations (visible as masked circles) are situated behind the frontier. Now, consider observation 1 denoted by ① in Figure 2. At its output level of 2, the intersection with a horizontal plane leads to the isoquant of an input set provided in Figure 1.

< FIGURE 2 ABOUT HERE >

To compute C^{CRS} for observation 1, start by computing $C1$ for all observations as indicated in Algorithm 1. For example, for the second observation this results in $C1 = \frac{2}{4} \cdot (4 \cdot 3 + 5 \cdot 2.5) = 12.25$. This value $C1$ is computed similarly for all observations leading to the values reported in the last column of Table 1. The smallest $C1$-value, that is 11.20, yields C^{CRS} for observation 1 and is indicated by the bordered cell in Table 1. This process can now be repeated for all other observations. In fact, it can also be executed for an arbitrary y-level.

5. CONCLUSION

This note is the first to prove that an enumeration algorithm can be employed to solve for certain specific convex DEA value-based models. Hitherto, enumeration has solely been applied to the specific structure of a non-convex production model (see, e.g., Ray (2004)). Apart from a general argument as to the computational complexity of enumeration versus LP, the empirical illustration reveals that potentially substantial computational gains are realistic. Obviously, in case bootstrapping techniques are used to conduct statistical inference (with a LP in each draw), the gains in computational burden will prove even more substantial. Perhaps, these gains are such as to justify inclusion of the above special cases in some of the dedicated DEA software around.

Obviously, we do not claim that enumeration is a viable solution strategy for convex DEA type of production- and value-based models in general. But, it cannot be excluded that enumeration could be applied to some other specific convex DEA models. For instance, to the extent that one is willing to select an efficiency measure that always projects onto a vertex point, the same procedure could probably be applied to production models under constant returns to scale and a single output or input with the measurement orientation along this single dimension (see, e.g., Russell and Schworm (2011) for some of the more recent choices). This could be a promising avenue for future research.

REFERENCES

Ali, I.A. (1993) Streamlined Computation for Data Envelopment Analysis, *European Journal of Operational Research*, 64(1), 61-67.

Atkinson, S.E, J.H. Dorfman, (2009) Feasible Estimation of Firm-Specific Allocative Inefficiency through Bayesian Numerical Methods, *Journal of Applied Econometrics*, 24(4), 675–697.

Banker, R., A. Charnes, W. Cooper (1984) Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science*, 30(9), 1078-1092.

Camanho, A.S., R.G. Dyson (2005) Cost Efficiency Measurement with Price Uncertainty: A DEA Application to Bank Branch Assessments, *European Journal of Operational Research*, 161(2), 432-446.

Charnes, A., W. Cooper, E. Rhodes (1978) Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research*, 2(6), 429-444.

Chong, E.K.P., S.H. Zak (2001) *Introduction to Optimization*, 2nd Ed, New York, Wiley.

Cooper, W.W., L.M. Seiford, K. Tone (2006) *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, 2nd Ed., Berlin, Springer.

Cooper, W.W., L.M. Seiford, J. Zhu (2011) Data Envelopment Analysis: History, Models, and Interpretations, in: W.W. Cooper, L.M. Seiford, J. Zhu (eds.) *Handbook on Data Envelopment Analysis*, 2nd Ed., Berlin, Springer, 1-39.

De Borger, B., K. Kerstens, M. Staat (2008) Transit Costs and Cost Efficiency: Bootstrapping Non-Parametric Frontiers, *Research in Transportation Economics*, 23(1), 53-64.

Dulá, J.H. (2002) Data Envelopment Analysis, in: P.M. Pardalos, M.G.C. Resende (eds.) *Handbook of Applied Optimization*, Oxford, Oxford University Press, 531-543.

Eiselt, H.A., C.-L. Sandblom (2007) *Linear Programming and its Applications*, Berlin, Springer.

Eremin, I.I. (2002) *Theory of Linear Optimization*, Utrecht, VSP.

Färe, R., S. Grosskopf, C.A.K. Lovell (1994) *Production Frontiers*, Cambridge, Cambridge University Press.

Fried, H.O., C.A.K. Lovell, S.S. Schmidt (2008) Efficiency and Productivity, in: H. Fried, C.A.K. Lovell, S. Schmidt (eds) *The Measurement of Productive Efficiency and Productivity Change*, New York, Oxford University Press, 3-91

Hackman, S.T. (2008) *Production Economics: Integrating the Microeconomic and Engineering Perspectives*, Berlin, Springer.

Jahanshahloo, G.R., M. Soleimani-damaneh, A. Mostafaee (2008) A Simplified Version of the DEA Cost Efficiency Model, *European Journal of Operational Research*, 184(2), 814-815.

Paryab, K., R. Khanjani Shiraz, L. Jalalzadeh (2012) An Improved Model for Assessing the FDH-Cost Efficiency, *Asia-Pacific Journal of Operational Research*, 29(4), 1250022.

Ray, S.C. (2004) *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research*, Cambridge, Cambridge University Press.

Russell, R.R., W. Schworm (2011) Properties of Inefficiency Indexes on $\langle$Input, Output$\rangle$ Space, *Journal of Productivity Analysis*, 36(2), 143-156.

Vassiloglou, M., D. Giokas (1990) A Study of the Relative Efficiency of Bank Branches: An Application of Data Envelopment Analysis, *Journal of the Operational Research Society*, 41(7), 591–597.

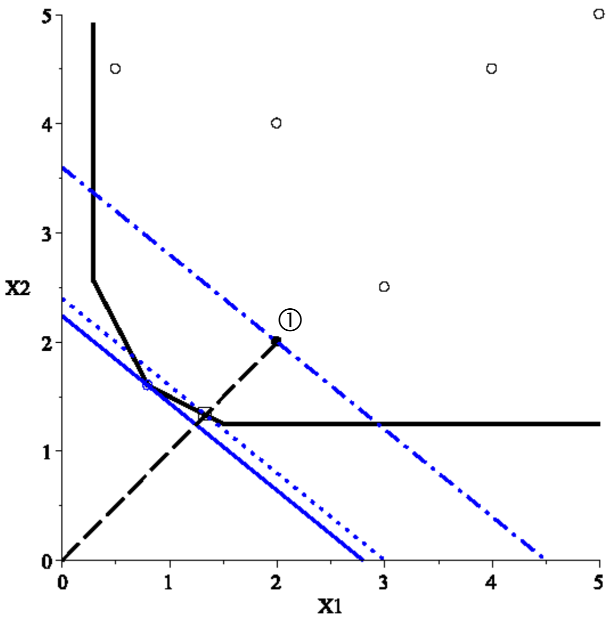

Figure 1: CRS production possibility set for observation 1 in Table 1 with visualization of the current cost level (dash-dotted line), technical efficient cost level (dotted line) and the minimal cost level (solid line).

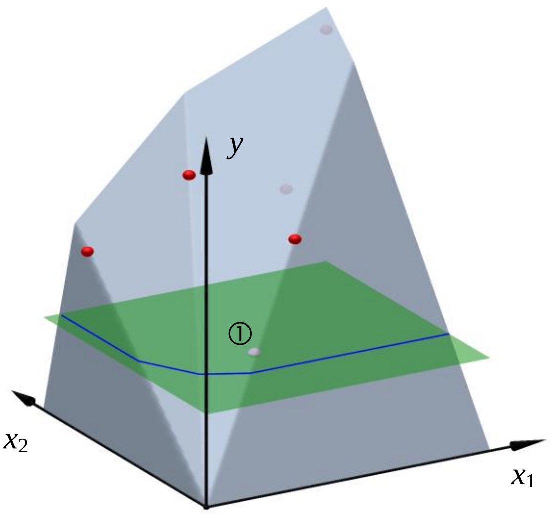

Figure 2: 3d-view of the observations given in Table 1, the CRS efficient frontier and its section by the horizontal plane at the output level of observation 1.