Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP04/2022

A Homothetic Data Generated Technology

Antonio Peyrache

Date: **May 2022**

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

A Homothetic Data Generated Technology

ANTONIO PEYRACHE*

a.peyrache@uq.edu.au

Abstract

In this paper I propose a method for constructing an enlargement of a variable returns to scale (VRS) data generated production technology that will satisfy homotheticity. The method can be used both with convex and non-convex technologies and both in the single output and multiple output setting. The method is computationally fast, therefore it provides a tool that can be used on large datasets. An empirical illustration is provided based on a dataset of Italian courts of justice.

Key Words: Input Homotheticity, Output Homotheticity, DEA, FDH, Efficiency

1 Introduction

In this paper I address the problem of building a data generated technology[1] that satisfies input (and/or output) homotheticity. In particular, I am interested in understanding if it is possible to define an appropriate enlargement of the VRS hull which is able to satisfy input (and/or output) homotheticity. I show that in the single output-multiple input (single input-multiple output) case it is possible to build such an enlargement easily by using the constant returns to scale (CRS) reference technology (the conical extension of the VRS hull). A similar

*Centre for Efficiency and Productivity Analysis (CEPA), School of Economics, The University of Queensland, Australia.

[1]Russell and Schworm (2009) called production sets that are based on axioms and use the minimum extrapolation principle (i.e. they contain a given set of observations) Data Generated Technologies. This includes both the data envelopment analysis (DEA) and the free disposal hull (FDH) models. I follow Russell and Schworm (2009) in calling these technologies Data Generated Technologies.

simple and direct procedure is not available in the general multiple output-multiple input case. In such a case I propose a method that provides an enlargement of the CRS cone that satisfies both input and output homotheticity and use it to provide an enlargement of the VRS hull that will satisfy input and/or output homotheticity. The method has the advantage of being easy to implement and computationally fast with similar computational time as standard convex (DEA) or non-convex (FDH) models. Moreover, the fact that a CRS reference set is used in building the homothetic extension of the VRS hull, implies that the conical extension of the VRS and homothetic VRS technologies coincide with the CRS technology.

Olesen (2014) introduced a method to find an enlargement of the VRS hull that satisfies homotheticity (the author extends an idea presented in Hanoch and Rothschild (1972)). The method proposed in Olesen (2014) is limited to the single output-multiple input case and the convex technology, while the method proposed here extends to multiple outputs technologies and non-convex technologies. Additionally, Olesen (2014) mentions potential computational issues in large datasets, a problem that does not arise with the model presented here. Moreover, if one is interested in imposing other regularity conditions such as the regular ultra passum law (see Olesen and Ruggiero (2014)), it is useful to have a method that can be used with non-convex technologies.

The simplicity and computational effectiveness of the method here proposed vindicates Ole Olesen original intuition that homotheticity can indeed be used as an additional axiom in building nonparametric production models (especially so in sinlge output technologies). Homotheticity may play an important role in allowing other properties of the graph of the technology to be imposed, such as direct estimation of economies of scale in the graph of the technology (Olesen and Ruggiero (2022)).

The method is illustrated by using data on courts of justice in Italy, first analyzed in Peyrache and Zago (2016). This illustration is intended to show in which way the method can be useful to the applied researcher (in particular how one can use it to obtain better estimates of input and output substitution possibilities).

The paper is organized as follows. In section 2 a numerical example is provided to show the main results and give an intuition of the proposed method. Section 3 introduces the notation and the definitions used. Section 4 describes the method for the single output-multiple input case. Section 5 extends to the case of multiple output-multiple input. Section 6 provides the empirical application and section 7 concludes.

2 Numerical Example

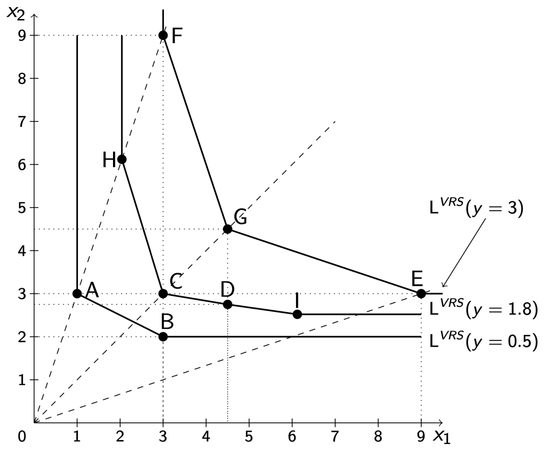
In table 1 I provide a numerical example of the proposed method to illustrate the basic concepts and strategy that are used in order to impose homotheticity onto the data. This is illustrated in the single output case, since this will provide the basic model and also because in this case there is a quite straightforward way of imposing homotheticity. There are 7 DMUs that are indicated by letters and produce three different output levels. Since there are only three output levels (0.5, 1.8, 3), there are associated with these levels three different input isoquants. The table also reports the ratio of inputs to output, the value of the aggregate input (described below) and the output oriented efficiency scores for the VRS, CRS and Homothetic-VRS (H-VRS) technologies.

DMU	Y	X_1	X_2	X_1/Y	X_2/Y	XAgg	VRS	CRS	H-VRS
A	0.5	1	3	2	6	1	1	0.5	1
B	0.5	3	2	6	4	1.5	1	0.33	0.43
C	1.8	3	3	1.67	1.67	2	1	0.9	1
D	1.8	4.5	2.75	2.78	1.53	2.13	1	0.85	0.92
E	3	9	3	3	1	3	1	1	1
F	3	3	9	1	3	3	1	1	1
G	3	4.5	4.5	1.5	1.5	3	1	1	1

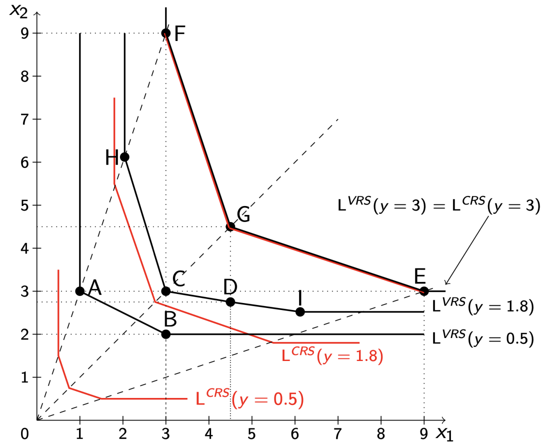
Table 1: Numerical Example: output oriented efficiency scores.

Under the assumptions of convexity and variable returns to scale (VRS) these three isoquants are shown in picture 1(a). Notice that two more points are included here: H and I. These are obtained as, respectively, the convex combinations: $(0.48A+0.52F)$ and $(0.48B+0.52E)$. These two points form part of the isoquant at the output level $y = 1.8$. The VRS isoquants are clearly non-homothetic (they are not "parallel", i.e. it is not possible to obtain the shape of one as a scaling of the shape of another one). Moreover all the 7 observations are lying on the frame of the VRS technology, therefore all of them are efficient. Each isoquant is indicated by the letter L, so $L^{VRS}(y = 0.5)$ will represent all the input combinations that are able to produce output level $y = 0.5$. In panel 1(b) both the VRS and the CRS isoquants are reported, with the CRS isoquants highlighted in red. Notice that the CRS technology is obtained as the conical extension of the VRS technology. Under the CRS assumption only three of the seven observations are efficient: E, F and G. For output level $y = 3$ the CRS and VRS isoquants coincides (this is not in general true, but it is for this example).

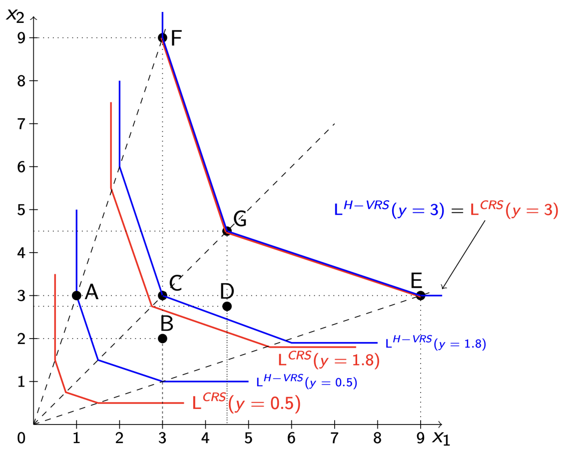
When scaling back the inputs and the output from the input isoquant $L^{CRS}(y = 3)$ we then obtain all the isoquants associated with the CRS cone technology. In particular the two that are of interest here are $L^{CRS}(y = 0.5)$ and $L^{CRS}(y = 1.8)$. Notice that the main characteristic of the red CRS isoquants is that now they are "parallel" shifts of each other, thus it is possible to obtain one isoquant as a scaling of the other one. This means that the CRS technology (only in the one output case) is also homothetic. Another feature of this numerical example is that observations A, B, C, D are lying in the non-decreasing returns to scale portion of the VRS technology. This is evidenced by the fact that the CRS technology makes use of less inputs to produce the same level of output compared to the VRS technology. It is now possible to use the fact that the CRS technology is homothetic to build an enlargement of the VRS technology that will satisfy homotheticity. This is done in panel 1(c) by considering in blue the homothetic VRS isoquants (H-VRS). This isoquants will have the same "shape" as the CRS ones, but will



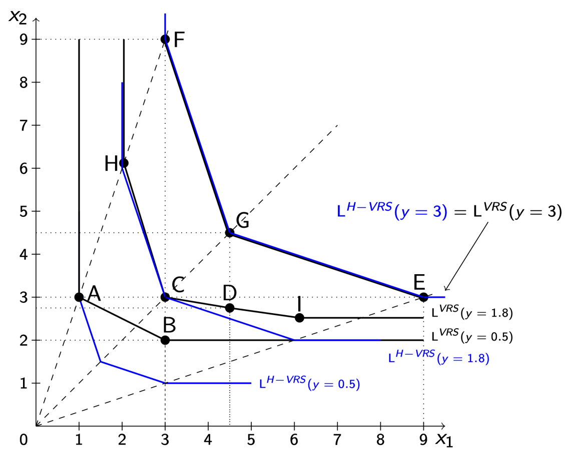
(a) VRS



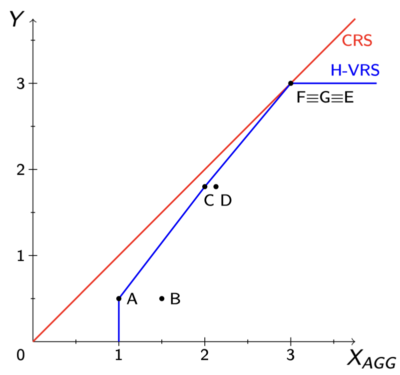
(b) VRS and CRS (in red)



(c) CRS (in red) and H-VRS (in blue)



(d) VRS and H-VRS (in blue)



(e) CRS and H-VRS in Aggregate Space

Figure 1: Input Isoquants for VRS, CRS and Homothetic VRS Technologies

lie at a different point on the graph. To obtain them, one has to "envelope" the data points under the constraint that the shape of the isoquant is given by the CRS isoquant. Therefore, for example, $L^{H-VRS}(y = 0.5)$ will have to envelope points A and B as tightly as possible. Clearly this is done in the picture and since the "shape" of the isoquant is given, DMU A will sit on the frame of this new technology and be efficient, but DMU B will sit inside the frame and be inefficient. A similar argument applies to the other two isoquants. In particular, since the isoquant $L^{CRS}(y = 3)$ is scale efficient, the H-VRS technology isoquant $L^{H-VRS}(y = 3)$ will coincide with it at the scale efficient size. It is important to notice here that if one were to take the conical extension of the H-VRS technology, then this would return the CRS technology. This is an important feature of this method, since it guarantees that the homothetic enlargement of the VRS technology is contained in the CRS technology and therefore the conical extensions of the VRS and H-VRS technologies coincide. In panel 1(d) the VRS and H-VRS technologies are contrasted. As it can be clearly seen from this picture, the H-VRS provides a minimal enlargement of the VRS technology that will satisfy homotheticity. DMUs B and D that were efficient under VRS, will now be inefficient if benchmarked against the H-VRS technology. This means that the H-VRS technology will have more discrimination power compared to the VRS technology, as expected. In the last panel 1(e) the technology is represented by using the input aggregates. To obtain these aggregates one picks one input isoquant of the H-VRS technology and then measures the distance of each DMU input vector from this isoquant. In this example, isoquant $L^{H-VRS}(y = 0.5)$ is chosen as a reference. Since DMU A lies on this isoquant the input aggregate for this DMU will be equal to $X_{Agg} = 1$. For DMU F, for example, the radial distance from this basic isoquant is OF/OA, therefore the input aggregate for DMU F is $X_{Agg} = 3$. The other DMUs input aggregates can be obtained similarly by measuring the radial distance from this base isoquant. It should also be noted that the choice of the base isoquant is innocuous, since picking a different isoquant will be equivalent to a unit of measurement change (multiplication by a scalar) of the input aggregates. In panel 1(e) all the combinations of input aggregates and output are plotted. The CRS and H-VRS technologies are represented here and the associated scale economies can easily be checked via this scatter plot. Notice that there is no such an aggregate space representation for the VRS technology, since in general the VRS technology will not satisfy input homotheticity and therefore it is deprived of an aggregate space representation.

It is interesting to note that the previous method also provides a way of determining the set of points that will constitute the frame of the homothetic VRS technology. Table 2 provides a reference set of observations that, if used, will generate a VRS hull that is input homothetic. If the efficiency of the original set of observations were to be measured against this new reference set, the efficiency score obtained for each observation would be the same as the one obtained in the aggregate space of figure 1(e). Moreover, the fact that a dataset satisfying homotheticity can be built, implies that further analysis of the data can be conducted directly on this dataset, therefore enforcing homotheticity on any subsequent result. It should also be noted here that imposing homotheticity is equivalent to building input aggregates. In other words, any set of input aggregates will imply an input homothetic structure and viceversa. Although this example has been carried out using the convexity assumption, nothing prevents from using a non-convex model. As one would expect, the convex homothetic VRS technology will be equal to the convex closure of the non-convex homothetic VRS technology, a fact that will be used later in section 5 in the building of a homothetic structure for the multiple output case.

DMU	Y	X_1	X_2
A’	0.5	1	3
B’	0.5	3	1
C’	0.5	1.5	1.5
D’	1.8	2	6
E’	1.8	3	3
F’	1.8	6	2
G’	3	9	3
H’	3	3	9
I’	3	4.5	4.5

Table 2: Observations that will Generate a Homothetic Frame

3 Definitions and Assumptions

Before extending the computational aspects described in the numerical example to the general case, it is important to recall some theoretical notions and definitions that will help and inform the subsequent computational strategy. To this purpose, consider a production process where inputs $\mathbf{x} \in \mathbb{R}_+^P$ are used to produce outputs $\mathbf{y} \in \mathbb{R}_+^Q$ and let T denote the production possibilities set or technology (see O'Donnell (2018)):

$$T = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}_+^P \times \mathbb{R}_+^Q \mid \mathbf{x} \text{ can produce } \mathbf{y}\}.$$

The output attainable set is the set of possible outputs $\mathbf{y}$ producible using input vector $\mathbf{x}$ and it can be defined as

$$P(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}_+^Q \mid (\mathbf{x}, \mathbf{y}) \in T\}.$$

The output isoquant is then defined as: $IP(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}_+^Q \mid (\mathbf{x}, \mathbf{y}/\theta) \notin T, \theta < 1\}$. Similarly, for any $\mathbf{y} \in \mathbb{R}_+^Q$ the input requirement set $L(\mathbf{y})$ can be defined as the set of all input vectors which yield at least $\mathbf{y}$:

$$L(\mathbf{y}) = \left\{\mathbf{x} \in \mathbb{R}_+^P \mid (\mathbf{x}, \mathbf{y}) \in T\right\}.$$

and the input isoquant is then defined as: $IL(\mathbf{y}) = \left\{\mathbf{x} \in \mathbb{R}_+^P \mid (\mathbf{x}/\lambda, \mathbf{y}) \notin T, \lambda > 1\right\}$. The conical extension of the production set T is defined as:

$$T^C = \{(\alpha\mathbf{x}, \alpha\mathbf{y}) : (\mathbf{x}, \mathbf{y}) \in T, \alpha \geq 0\}$$

The technology satisfies CRS if and only if $T = T^C$. Note that, in general, even if T does not satisfy CRS, the conical extension T^C can be defined as an enalrgement of T according to the previous formula. This will imply some restrictions on the production set T in order for T^C to be a strict subset of the positive orthant[2].

3.1 Data Generated Technologies

In order to give empirical meaning to the previous definitions consider the production possibilities set generated by a set of data points or decision making units (DMUs) $(\mathbf{x}_k, \mathbf{y_k})$ $(\forall k = 1, \ldots, K)$, where the input and output vectors are column vectors. In the spirit of Podinovski (2022), I make the following assumptions on the technology set:

A1. Feasibility of observed DMUs: the observed data points belong to the production set $((\mathbf{x}_k, \mathbf{y}_k) \in T)$.

A2. Free Disposability: $\forall(\mathbf{x}, \mathbf{y}) \in T$, if $\mathbf{x}' \geq \mathbf{x}$, $\mathbf{y}' \leq \mathbf{y}$, then $(\mathbf{x}', \mathbf{y}') \in T$.

A3. Convexity: $(\mathbf{x}_1, \mathbf{y}_1) \in T$ and $(\mathbf{x}_2, \mathbf{y}_2) \in T \implies [\gamma \mathbf{x}_1 + (1-\gamma)\mathbf{x}_2, \gamma \mathbf{y}_1 + (1-\gamma)\mathbf{y}_2] \in T$, $0 \leq \gamma \leq 1$.

Notice that this minimal set of axioms implies that $L(\mathbf{y})$ is closed and $P(\mathbf{x})$ is compact. The convexity assumption is not necessary in any of the results presented below, but it is here introduced because of its widespread use. One can apply the method proposed in this study to both convex and non-convex settings, therefore there is no harm in dispensing of this assumption. The numerical example above makes use of convexity for illustrative purposes and in the empirical computations below both the convex and the non-convex technology results will be presented. Notice that homotheticity and convexity are two separate properties and one should, ideally, be able to impose them separately on the technology (as I shall propose in this paper). The first two assumptions are, on the contrary, maintained throughout, although one may relax free disposability and use the weak disposability axiom (this would perhaps involve

[2] Since the technology is generated by a set of given data points, what is required is that there are no observations that produce strictly positive output using a zero input vector. This is also known as the no free lunch assumption: $\nexists k :\ \mathbf{x}_k = \mathbf{0}_P \, and \, \mathbf{y}_k > 0$.

maintaining convexity to obtain sensible results). The production technology implied by this set of axioms in conjunction with the observed data points is:

$$T = \left\{ (\mathbf{x}, \mathbf{y}) : \sum_k \lambda_k \mathbf{x}_k \leq \mathbf{x}, \sum_k \lambda_k \mathbf{y}_k \geq \mathbf{y}, \sum_k \lambda_k = 1 \right\} \tag{1}$$

This technology is known as the VRS technology and it is possible to add the binary constraint ($\lambda_k \in \{0, 1\}$) to allow for a non-convex technology. The non-convex technology is known as the free disposal hull (FDH) technology. Since the convexity assumption can be dispensed of without affecting any of the results, I will refer simply to T as the data generated technology under VRS and this should be intended as including both the convex and non-convex case. The CRS conical extension of this set is:

$$T^C = \left\{ (\mathbf{x}, \mathbf{y}) : \sum_k \lambda_k \mathbf{x}_k \leq \delta \mathbf{x}, \sum_k \gamma_k \mathbf{y}_k \geq \delta \mathbf{y}, \sum_k \lambda_k = 1, \delta \geq 0 \right\} \tag{2}$$

It should be noted that under the binary constraint ($\lambda_k \in \{0, 1\}$), the conical extension of the set is a non-convex cone. This non-convex cone technology has been studied in Kerstens and Eeckaut (1999), Podinovski (2004). Briec and Kerstens (2006) show that computation of distance functions using a non-convex cone technology can be accomplished by using an enumeration algorithm, as opposed to a linear program (LP) for the convex technology. This makes computation over large datasets for the non-convex case feasible and fast; in fact, orders of magnitude faster than in the convex case.

3.2 Distance Functions, Efficiency and Homotheticity

The technical efficiency of any given point $(\mathbf{x}, \mathbf{y}) \in T$ is determined by the distance of the point to the boundary of the attainable set. Input and output distance functions measure the technical efficiency in the input and output direction respectively and are defined as:

$$D_I(\mathbf{x},\mathbf{y}) = \sup\{\lambda > 0 \mid (\mathbf{x}/\lambda, \mathbf{y}) \in T\} \tag{3}$$

and

$$D_O(\mathbf{x},\mathbf{y}) = \inf\{\theta > 0 \mid (\mathbf{x}, \mathbf{y}/\theta) \in T\}. \tag{4}$$

Under our assumptions on the data generated technology, the input and the output distance functions will satisfy a number of properties. For the results presented in this paper, it is important to point out that, by definition, the input distance function is linearly homogeneous in the input vector and the output distance function is linearly homogeneous in the output vector. For any given dataset, these quantities can be therefore computed as a linear program in the case of a convex technology and using an enumeration algorithm in the case of a non-convex technology.

We say that the technology exhibits output homotheticity if the output set satisfies $P(\mathbf{x}) = P(\mathbf{1})G(\mathbf{x})$, with $G(\mathbf{x})$ being a non-decreasing function in its argument; this implies the following functional separability of the output distance function:

$$D_O(\mathbf{x},\mathbf{y}) = \frac{Y(\mathbf{y})}{G(\mathbf{x})} \tag{5}$$

where $Y(\mathbf{y})$ is a linearly homogeneous, non-decreasing function in the output vector. This function aggregates the output vector into a single index. Notice that this simple fact implies that the technology can be represented using a production function: $Y(\mathbf{y}) = \theta G(\mathbf{x})$, with $\theta = D_O(\mathbf{x},\mathbf{y})$ being the efficiency of production. This means that output homotheticity allows to separate functionally the outputs from the inputs and express every output vector as a single valued aggregate number. For example if for two output vectors $\mathbf{y}_1, \mathbf{y}_2$, $Y(\mathbf{y}_1) = Y(\mathbf{y}_1)$, this means that the two vectors are of the same "size" although they may have a different composition: the two vectors lie on the same output isoquant if and only if the two aggregates have the same value. In other words, the notion of a given "level" or "size" of output production takes a very transparent meaning under output homotheticity, something that is not possible with multiple outputs under non-homothetic technologies. The output aggreagates $Y(\mathbf{y})$ can be build, in practice, by choosing a reference input vector, say $\mathbf{x} = \mathbf{1}_P$:

$$Y(\mathbf{y}_k) = D_O(\mathbf{1}_P, \mathbf{y}_k) \tag{6}$$

where it is clear that due to the separability of the output distance function, the choice of the reference input vector $\mathbf{x} = \mathbf{1}_P$ is innocuous, since any other reference input vector would give rise to the same output aggregates (up to a re-scaling factor). Thus the output aggregates are uniquely determined up to an innocuous scaling normalization that amounts to the equivalent of a choice on the unit of measuremet of the output aggregate[3]. Associated with the output homothetic structure, I define an alternative representation of the technology based directly on the output aggregates:

$$T^{AOH} = \left\{ (\mathbf{x}, Y) : \sum_k \lambda_k \mathbf{x}_k \leq \mathbf{x}, \sum_k \lambda_k Y_k \geq Y, \sum_k \lambda_k = 1 \right\} \tag{8}$$

where $Y_k = Y(\mathbf{y}_k)$ are the output aggregates computed above. One can also consider the conical extension of this aggregate technology T^{AOHC}. Because of the separability of the output distance function one can measure efficiency against this aggregate technology, basically reducing it to a single output technology.

[3]An alternative approach for the construction of the output aggregates is to choose a different reference input vector and then measure the distance of each output vector with respect to this isoquant. For example, choosing $P(\mathbf{x}_j)$ will return the following output aggregates:

$$Y(\mathbf{y}_k) = \frac{D_O(\mathbf{x}_j, \mathbf{y}_k)}{D_O(\mathbf{x}_j, \mathbf{y}_1)} \tag{7}$$

This is also known as the Malmquist output quantity index. It is worth stressing again that these output aggregates are the same as the ones obtained before (under output homotheticity), except for a re-scaling. In other words the ratio between any two output aggregates is the same using the two methods.

The technology satisfies input homotheticity if the input set satisfies $L(\mathbf{y}) = L(\mathbf{1})H(\mathbf{y})$, with $H(\mathbf{y})$ a non-decreasing function of its argument; this implies the following functional separation of the input distance function:

$$D_I(\mathbf{x}, \mathbf{y}) = \frac{X(\mathbf{x})}{H(\mathbf{y})} \tag{9}$$

where $X(\mathbf{x})$ is a linearly homogeneous, non-decreasing function in the input vector. A similar interpretation holds here for the input aggregate $X(\mathbf{x})$ as discussed previously for the output aggregate $Y(\mathbf{y})$. In particular by choosing as a reference the unit output vector $\mathbf{y} = \mathbf{1}_Q$, the input aggregates can be obtained as:

$$X(\mathbf{x}_k) = D_I(\mathbf{x}_k, \mathbf{1}_Q) \tag{10}$$

where due to the separability of the input distance function, the choice of the reference output vector is innocuous, since any other reference output vector would give rise to the same input aggregates (up to a re-scaling factor)[4]. Associated with the input homothetic structure, I define an alternative representation of the technology based directly on the input aggregates:

$$T^{AIH} = \left\{ (X, \mathbf{y}_k) : \sum_k \lambda_k X_k \leq X, \sum_k \gamma_k \mathbf{y}_k \geq \mathbf{y}, \sum_k \lambda_k = 1 \right\} \tag{12}$$

where $X_k = X(\mathbf{x}_k)$ are the input aggregates. One can also consider the conical extension of this aggregate technology T^{IAHC}. Because of the separability of the input distance function, one can measure efficiency against this aggregate technology.

A technology which jointly satisfies input and output homotheticity has the following rep-

[4]Similarly to the output case, the input aggregates can also be computed as:

$$X(\mathbf{x}_k) = \frac{D_I(\mathbf{x}_k, \mathbf{y}_j)}{D_I(\mathbf{x}_1, \mathbf{y}_j)} \tag{11}$$

This is known as the Malmquist input quantity index. Again, we are choosing just a different scaling of the input aggregates, leaving their relative ratios constant.

resentation in terms of the output distance function:

$$D_O(\mathbf{x}, \mathbf{y}) = \frac{Y(\mathbf{y})}{F[X(\mathbf{x})]} \tag{13}$$

where $F(\cdot)$ is a non-decreasing function of its scalar argument. Joint input and output homotheticity can also be represented functionally using the input distance function:

$$D_I(\mathbf{x}, \mathbf{y}) = \frac{X(\mathbf{x})}{G[Y(\mathbf{y})]} \tag{14}$$

where $G(\cdot)$ is a non-increasing function of its scalar argument. Associated with the joint homothetic structure, I define an alternative representation of the technology based directly on the output and input aggregates:

$$T^{AH} = \left\{ (X, Y) : \sum_k \lambda_k X_k \leq X, \sum_k \gamma_k Y_k \geq Y, \sum_k \lambda_k = 1 \right\} \tag{15}$$

The previous definitions are applied to T, the VRS hull. It is important to explore how these notions apply to the conical extesion T^C of the VRS set. The input distance function associated with the conical extension is defined as

$$D_I^C(\mathbf{x}, \mathbf{y}) = \sup \left\{ \lambda > 0 \mid (\mathbf{x}/\lambda, \mathbf{y}) \in T^C \right\} \tag{16}$$

and it will satisfy linear homogeneity in $\mathbf{x}$ and homogeneity of degree -1 in $\mathbf{y}$. The output distance function associated with the conical extension is defined as

$$D_O^C(\mathbf{x}, \mathbf{y}) = \inf \left\{ \theta > 0 \mid (\mathbf{x}, \mathbf{y}/\theta) \in T^C \right\}. \tag{17}$$

and it satisfies linear homogeneity in $\mathbf{y}$ and homogeneity of degree -1 in $\mathbf{x}$. The only difference here is that the distance functions are defined over the set T^C instead of the set T. Moreover for the conical extension T^C, $D_O^C(\mathbf{x},\mathbf{y}) = 1/D_I^C(\mathbf{x},\mathbf{y})$. Clearly, if the production set T is output homothethic, the conical extension T^C will inherit output homotheticity and the output distance function associated with the conical extension will satisfy the following functional restriction:

$$D_O^C(\mathbf{x},\mathbf{y}) = \frac{Y(\mathbf{y})}{G^C(\mathbf{x})} \tag{18}$$

where $Y(\mathbf{y})$ is the same as in equation (5), but $G^C(\mathbf{x})$ is now linearly homogeneous in its argument. If the production set satisfies input homotheticity, the input distance function associated with the conical extension will satisfy the following functional restriction:

$$D_I^C(\mathbf{x},\mathbf{y}) = \frac{X(\mathbf{x})}{H^C(\mathbf{y})} \tag{19}$$

where $X(\mathbf{x})$ is the same as in equation (9), but $H^C(\mathbf{y})$ is linearly homogeneous in its argument. This also means that for the conical extension T^C, $H^C(\mathbf{y}) = a_y Y(\mathbf{y})$ and $G^C(\mathbf{x}) = a_x X(\mathbf{x})$, i.e. these functions are linear functions in their respective aggregates (after appropriate normalization) and $a_y = a_x^{-1}$. One last property associated with the conical extension of the technology is important for this discussion. For a single output ($y \in \mathbb{R}_+$) multiple input ($\mathbf{x} \in \mathbb{R}_+^p$) technology, CRS is sufficient for input homotheticity (the reverse is not true and a technology can be input homothetic without being CRS; this is trivial, otherwise the *raison d'être* of this paper would be undermined). Similarly, for a single input ($x \in \mathbb{R}_+$) multiple output ($\mathbf{y} \in \mathbb{R}_+^q$) technology, CRS is sufficient for output homotheticity (the reverse is not true). This result points to the fact that a simple way of imposing homotheticity for a single output (or single input) data generated technology is to look at its conical extension: the conical extension will always satisfy CRS and consequently homotheticity. This result will be used in the next section to build a homothetic VRS hull. Another fact should be noted here. If a technology satisfies CRS then input homotheticity will imply output homotheticity and viceversa. In other words, for the conical extension, input homotheticity will hold if and only if output homotheticity holds. This fact implies that:

$$D_O^C(\mathbf{x}, \mathbf{y}) = \frac{Y(\mathbf{y})}{G[X(\mathbf{x})]} \Leftrightarrow D_I^C(\mathbf{x}, \mathbf{y}) = \frac{X(\mathbf{x})}{H[Y(\mathbf{y})]} \tag{20}$$

and this relationship will be used to deliver a method for imposing homotheticity in the multiple input-multiple output case. These relationships will be useful in order to inform our strategy on how to impose homotheticity on an otherwise non-homothetic technology.

In general for any given set of data points $(\mathbf{x}_k, \mathbf{y}_k)$, the data generated technology under VRS will not satisfy homotheticity. The question then arises if it is possible to build an enlargement of this technology that satisfies homotheticity as an additional production axiom. It is important to stress that, since the VRS data generated technology does not in general satisfy homotheticity, the separability conditions on the distance functions discussed in this section will in general not hold for any particular dataset and its associated data generated technology. The problem then becomes how to use the theoretical relationships in order to build an enlargement of the production technology that will satisfy homotheticity, therefore the separability conditions on the distance functions. The next sections should address such a problem.

4 Single output case

In this section I consider the problem of imposing an input homothetic structure on the VRS technology (1) for the single output case. The method works in a similar way for the single input multiple output case. In order to build the input aggregates implied by the input homothetic structure, the unit input isoquant of the CRS technology T^C is chosen as a reference $L^C(y = 1)$. This means solving the program associated with the computation of the following input distance function $X(\mathbf{x}_k) = D_I^C(\mathbf{x}_k, 1)$ for all observations in the dataset. For the sake of clarity, the program is reported here for a generic observation $o = 1, \dots, K$:

$$\begin{aligned} \min \quad & \theta \\ st \quad & \textstyle\sum_k \lambda_k \mathbf{x}_k \leq \theta \mathbf{x}_o \\ & \textstyle\sum_k \lambda_k y_k \geq 1 \end{aligned} \tag{21}$$

where the optimal value of the objective function will return the input aggregate: $X(\mathbf{x}_o) = 1/\theta^*$. To obtain the program for the non-convex case, one adds the following two constraints: $\lambda_k \leq D\beta_k$ and $\sum_k \beta_k = 1$, where D is an appropriate large number and $\beta_k \in \{0, 1\}$ are binary variables. It is interesting to note that by a transformation of variables ($\mu_k = \lambda_k y_k$), this program can be written as:

$$\begin{aligned} \min \quad & \theta \\ st \quad & \textstyle\sum_k \mu_k \mathbf{a}_k \leq \theta \mathbf{x}_o \\ & \textstyle\sum_k \mu_k \geq 1 \end{aligned} \tag{22}$$

where $\mathbf{a}_k = \dfrac{\mathbf{x}_k}{y_k}$ and this highlights that the base isoquant can be obtained by expressing the variables in intensive form $(\mathbf{x}/y)$. This will provide the reference base isoquant for the building of the aggregators, and it vindicates the intuition presented in the numerical example of section 2. Given the $X_k = X(\mathbf{x}_k)$ input aggregates, the output oriented efficiency scores under the VRS input homothetic technology can be computed using the aggregate input homothetic technology (T^{AIH}) defined in equation (12) (one can also compute the input oriented or any other type of orientation):

$$D_O^H(\mathbf{x}_k, y_k) = D_O^{AIH}(X_k, y_k) \tag{23}$$

where $D_O^H(\mathbf{x}_k, y_k) \leq D_O(\mathbf{x}_k, y_k)$. The implied program is:

$$
\begin{aligned}
\max \quad & \theta \\
st \quad & \textstyle\sum_k \lambda_k X_k \leq X_o \\
& \textstyle\sum_k \lambda_k y_k \geq \theta y_o \\
& \textstyle\sum_k \lambda_k = 1
\end{aligned}
\tag{24}
$$

and $D_O^H(\mathbf{x}_k, y_k) = 1/\theta^*$. Clearly, the homothetic technology is a single input-single output technology and the aggregator function $X(\mathbf{x})$ describes the input substitution possibilities of the technology. There are various ways of determining the frame of the homothetic VRS technology in the full input-output space, as opposed to the aggregate space. One (very intuitive?) way is to look at the following enlarged dataset:

$$
\left(\mathbf{x}_k \frac{X_j}{X_k}, y_j\right), \forall k, j = 1, \ldots, K \tag{25}
$$

The data generated technology (1) obtained with this enlarged dataset will return a homothetic frame as an enlargement of the VRS technology. Since the input aggregates have been determined already, the construction of this enlarged dataset does not present any particular computational burden. The efficiency scores based on this enlarged technology under VRS will be the same as the ones in equation (23). The problem with this description of the frame is that it is not parsimonious, requiring K^2 data points, therefore making computation of efficiency scores slow. A much better way of building the frame is to list the observations that form the base unit isoquant, these can be obtained as the set of efficient observations when efficiency is benchmarked against the CRS technology. In other words, these are the observations that form the frame of the CRS technology. Collect these $N \leq K$ efficient points in the matrix $\mathbf{A} = [\mathbf{a}_n]$ for the unit output level $y = 1$ (with variables expressed in intensive form, as ratios to their own output level: $\mathbf{a}_n = \mathbf{x}_n / y_n$). This matrix is of dimension $P \times N$ and has on the columns the input vectors that compose the base unit isoquant. Notice, moreover, that one does not need all the observations to describe the homothetic frame (as much as one does not need all observations to describe the CRS or VRS frame), but only observations that are efficient when benchmarked against the homothetic VRS technology (this can be determined by looking at the efficiency scores determined in equation (23)). If this number is $M < K$, then the description of the homothetic frame can be done using only these M efficient DMUs. Then the following dataset will return a parsimonious way of determining the frame of the homothetic VRS technology:

$$\begin{bmatrix} \mathbf{Z} \\ \\ \mathbf{V} \end{bmatrix} = \begin{bmatrix} X_1\mathbf{A} & \dots & X_M\mathbf{A} \\ \\ y_1\mathbf{1}_N & \dots & y_M\mathbf{1}_N \end{bmatrix} \tag{26}$$

where matrices $\mathbf{Z}$ and $\mathbf{V}$ are, respectively, of dimension $P \times J$ and $Q \times J$ $(1 \times J)$, with $J = N \times M$ being the number of observations needed to build the homothetic frame. Notice that here N is the number of points needed to build the base isoquant and M the number of points needed to shift the base isoquant to the appropriate output level. For example in the numerical example presented in section 2, the first description of the frame will return a dataset of 49 observation, but only 9 are necessary to build the homothetic frame: $N = 3$ points that form the base isoquant and $M = 3$ different output levels. The homothetic VRS technology can then be obtained as follows:

$$T^H = \left\{ (\mathbf{x}, \mathbf{y}) : \sum_j \lambda_j \mathbf{z}_j \leq \mathbf{x}, \sum_j \lambda_j \mathbf{v}_j \geq \mathbf{y}, \sum_j \lambda_j = 1 \right\} \tag{27}$$

where $(\mathbf{z}_j, \mathbf{v}_j)$ are the columns of the associated matrices. The use of this technology will return a separable input distance function. Since the whole method can be used both in the convex and non-convex case, the convex homothetic technology will be equal to the convex closure of the non-convex homothetic technology. This concludes the construction of a homothetic VRS technology in the single output case (extension to the single input-multiple output case is trivial, following the same line of reasoning).

5 Multiple output case

In the multiple input-multiple output case the strategy outlined above will not work, since the conical closure of T, T^C, is in general neither input nor output homothetic (the CRS cone will be ray-homothetic, but this is of little help in this discussion). Neverthless, by recalling the fact that output homotheticity will imply input homotheticity (and viceversa) for a CRS technology, it is possible to reduce the problem to the one of imposing input or output homotheticity (since this will entail an output or input homothetic structure for the cone CRS technology). In the multiple output context, since the CRS cone is not homothetic, one needs to look for an enlargement of the cone itself that will then satisfy homotheticity. Since homotheticity is implied by a set of input or output aggregates, this means that a strategy must be deployed to compute either the output aggregator function $Y(\mathbf{y})$ or the input aggregator function $X(\mathbf{x})$ before proceeding to look at imposing joint homotheticity. The basic intuition here is that the conical closure of the technology must be enlarged in such a way as to produce a homothetic cone (the term "homothetic cone" is here used to mean that the CRS cone technology is both input and output homothetic, thus joint homothetic). This will be accomplished with an iterative procedure.

The input aggregates are obtained in the first step using equation (11) like in the previous section. This uses the conical CRS technology as a benchmark: $X(\mathbf{x}_k) = D_I^C(\mathbf{x}_k, \mathbf{1}_Q)$. Contrary to the single output case, in the multiple output case the choice of the reference output vector will return alternative input aggregates even under CRS, since the CRS cone is neither input nor output homothetic in this case[5]. The output aggregates are obtained, similarly, by computing:

[5]Thus, in general $\frac{D_I^C(\mathbf{x}_k,\mathbf{1})}{D_I^C(\mathbf{x}_1,\mathbf{1})} \neq \frac{D_I^C(\mathbf{x}_k,\mathbf{y}_j)}{D_I^C(\mathbf{x}_1,\mathbf{y}_j)}$ and the choice of alternative reference output vectors $\mathbf{y}_j$ will return different input aggregates. The difference between the two aggregates comes from the different composition (mix) of the outputs, not the difference in their "level" or "magnitude". Since the CRS technology is used as a reference, ray-homotheticity makes the two sets of aggregates equal under a re-scaling of the reference output vector: $\frac{D_I^C(\mathbf{x}_k,\mathbf{y}_j)}{D_I^C(\mathbf{x}_1,\mathbf{y}_j)} = \frac{D_I^C(\mathbf{x}_k,\alpha\mathbf{y}_j)}{D_I^C(\mathbf{x}_1,\alpha\mathbf{y}_j)}$, $\alpha > 0$. To this purpose, in practice, it is appropriate and recommended the use of the average output vector in the sample as a reference. To simplify notation, I will use the unit output vector and assume that the dataset has been normalized by the data mean, therefore returning an average vector of ones.

$Y(\mathbf{y}_k) = D_O^C(\mathbf{1}_P, \mathbf{y}_k)$. These output aggregates are also dependent on the choice of the reference input set.

Notice that because of the theoretical discussion above, the conical extension of the output homothetic structure induced by the output aggregates will implicitly define a set of input aggregates (and viceversa). This input aggregates will in general be different from the input aggregates obtained above. Another way of looking at this, is to say that the CRS cone implied by the output homothetic structure induced by the aggregates $Y(\mathbf{y}_k) = D_O^C(\mathbf{1}_P, \mathbf{y}_k)$ is not the same cone as the one implied by the input homothetic structure induced by the input aggregates $X(\mathbf{x}_k) = D_I^C(\mathbf{x}_k, \mathbf{1}_P)$. This is in contrast to the above theoretical result that states that the two structures should give rise to the same conical extension. Call T^{OH} the output homothetic technology under VRS implied by the output aggregates, T^{IH} the input homothetic technology under VRS implied by the input aggregates and T^{OHC}, T^{IHC} the conical extensions of these two sets. Then as established in section 3, $T^{OHC} = T^{IHC} = T^{HC}$, but in general these cones will differ if the aggregates are computed with the method above, in contrast to the theory. To avoid such an inconsistency an iterative procedure is deployed below. Before doing so, it is necessary to define the frames of the T^{IH} and T^{OH} technologies. The input base isoquant is determined as the set of observations that are efficient when benchmarked against the conical extension of technology (8), the output homothetic aggregate technology T^{AOH}. Say the number of these observations is N, collected in the matrix $\mathbf{A} = [\mathbf{a}_n]$, then the frame of the input homothetic technology is given by:

$$\begin{bmatrix} \mathbf{Z} \\ \mathbf{V} \end{bmatrix} = \begin{bmatrix} X_1\mathbf{A} & \dots & X_M\mathbf{A} \\ \mathbf{y}_1\mathbf{1}_N & \dots & \mathbf{y}_M\mathbf{1}_N \end{bmatrix} \tag{28}$$

Notice that the main difference with respect to the frame in the single output case comes from the fact that vectors instead of scalars are used for the output. This frame can be used to build the input homothetic technology T^{IH} using definition (27). The frame of the input homothetic technology T^{IH} can be used to build the input aggregates using its conical extension. These aggreagtes will be the same as the ones obtained using the aggregate output homothetic technology T^{AOH} conical extension, but different from the ones computed above: $D_I^{IHC}(\mathbf{x}_k, \mathbf{1}) = D_I^{AOHC}(\mathbf{x}_k, 1) \neq D_I^C(\mathbf{x}_k, \mathbf{1}_Q)$.

The output base isoquant is determined as the set of observations that are efficient when benchmarked against the conical extension of technology (12), the input homothetic aggregate technology T^{AIH}. Say the number of these observations is L, collected in the matrix $\mathbf{B} = [\mathbf{b}_l]$, then the frame of the output homothetic technology is given by:

$$\begin{bmatrix} \mathbf{Z} \\ \\ \mathbf{V} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \mathbf{1}_L & \dots & \mathbf{x}_M \mathbf{1}_L \\ \\ Y_1 \mathbf{B} & \dots & Y_M \mathbf{B} \end{bmatrix} \tag{29}$$

This frame can be used to build the output homothetic technology using definition (27). The frame of the output homothetic technology T^{OH} can be used to build the output aggregates using its conical extension. These aggreagtes will be the same as the ones obtained using the aggregate input homothetic technology T^{AIH} conical extension but different from the ones computed above: $D_O^{OHC}(\mathbf{1}, \mathbf{y}_k) = D_O^{AIHC}(1, \mathbf{y}_k) \neq D_O^C(\mathbf{1}_P, \mathbf{y}_k)$.

There are two main problems with this initial step and both of them suggest a way of building the iteration steps below. First, the conical extensions of T^{IO} and T^{OH} should be the same, but those are in general different after this initial step. Second, if the convex and non-convex models are applied separately, the frame of the convex technology is not necessarily contained in the frame of the non-convex technology. In other words, the convex technology may not be equal to the convex closure of the non-convex technology, and the two may in fact intersect. In order to obtain a unique conical extension of the two technologies and to obtain the convex sets as a convex closure of the non-convex sets, the following iterations can be implemented. Call the value of the input and output aggregates at step s of the iteration X_k^s and Y_k^s respectively (using as starting values the aggregates just computed). Then, at step $s \geq 2$:

1. compute both the convex and non-convex frames (using the convex and non-convex aggregates from the previous step) of technologies T^{IH} and T^{OH}; call T_C^{IH} and T_{NC}^{IH} the convex and non-convex sets respectively for the input homothetic structure (and similarly T_C^{OH} and T_{NC}^{OH} for the output homothetic structure); take the union of the convex and non-convex frames by stacking the respective matrices of data points; call these extended sets $T^{EIH} = T_C^{IH} \cup T_{NC}^{IH}$ and $T^{EOH} = T_C^{OH} \cup T_{NC}^{OH}$;

2. compute the convex and non-convex input aggregates using the conical extension of T^{EOH} as a reference: $X_k^s = D_I^{EOHC}(\mathbf{x}_k, \mathbf{1})$;

3. compute the convex and non-convex output aggregates using the conical extension of T^{EIH} as a reference: $Y_k^s = D_O^{EIHC}(\mathbf{1}, \mathbf{y}_k)$;

4. Repeat steps 1), 2) and 3) until a stopping criterion is reached, and the four sets of aggregates converge[6] (both for the convex and non-convex quantities).

The previous iteration procedure provides a way of computing the four sets of aggregates consistently with each other[7]. This means that the CRS cone implied by the input aggregates will be the same as the CRS cone implied by the output aggregates ($T^{OHC} = T^{IHC} = T^{HC}$). Moreover, this means that the convex homothetic technology will always be the convex closure of the homothetic non-convex technology, therefore respecting the principle that the two are connected. As a consequence, the conical extension of the convex technology will be equal to

[6]A stopping criterion can be defined by setting tolerance levels on the changes $\max_k(|Y_k^s - Y_k^{s-1}|)$ and $\max_k(|X_k^s - X_k^{s-1}|)$.

[7]This iterative procedure is relatively fast. In the empirical example below, it took about 3 minutes to converge to a tolerance level set at the 6th decimal digit. The biggest burden in the computational speed is given by the convex linear programs. One could take the view that it is unnecessary to run the convex models, by relying only on the non-convex model and then defining the convex model on the frame of the non-convex model. This would reduce the computational burden drastically, since enumeration can be used on all non-convex models. This is likely to take a fraction of a second even on large datasets.

the convex closure of the conical extension of the non-convex technology. It should be noted that all these relationships hold in the single output case, therefore the iteration procedure will just be a mean to preserve these theoretical relationships in the multiple output case. An additional interesting fact is that the convex input aggregates will always be larger or equal to the non-convex input aggregates and the convex output aggregates will always be smaller or equal to the convex output aggregates.

The efficiency scores under output, input or joint homotheticity can be determined respectively by using the aggregate technologies T^{AOH}, T^{AIH} and T^{AH} reported in equations (8), (12) and (15), together with the set of input and output aggregates determined by the previous iterative procedure. This reduces to computing efficiency scores for: a single output-multiple input technology in the case of output homotheticity (technology 8); a multiple output-single input technology for the case of input homotheticity (technology 12); and a single output-single input technology for joint homotheticity (technology 15). The frame of these three technologies can be obtained by determining the set of observations that form the input and output base isoquants. The frames for the input and output homothetic case have been reported above already. The frame of the joint homothetic technology is determined as:

$$\begin{bmatrix} \mathbf{Z}_m \\ \mathbf{V}_m \end{bmatrix} = \begin{bmatrix} X_m\mathbf{A} & \dots & X_m\mathbf{A} \\ Y_m\mathbf{1}_N\mathbf{b}_1 & \dots & Y_m\mathbf{1}_N\mathbf{b}_L \end{bmatrix}, \ m = 1, \dots, M \tag{30}$$

and then stacking these matrices one obtains the frame:

$$\begin{bmatrix} \mathbf{Z} \\ \mathbf{V} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 & \dots & \mathbf{Z}_M \\ \mathbf{V}_1 & \dots & \mathbf{V}_M \end{bmatrix} \tag{31}$$

The joint homothetic technology T^H can be built using this frame as a reference in definition (27). Each of these three technologies has a conical extension. Call T^{HC}, T^{IHC}, T^{OHC} these conical extensions. Then $T^{HC} = T^{IHC} = T^{OHC}$ as implied by the theory summarized above. This means that the conical extensions of these three technologies coincide. This concludes the discussion of the multiple output case. One should note that the frame of the various sets is determined in a very parsimonious way, since in the worst case scenario (joint homotheticity) one is dealing with $J = N \times L \times M$ points: N points will describe the input isoquant, L points will describe the output isoquant and M points represent the different levels of production. This number is likely to be much smaller, in practice, than the original dataset with K observations (although one cannot exclude that this set of observations may actually be bigger).

6 Empirical application

In this section a real data example is used to illustrate the practical use of the homotheticity assumption. Peyrache and Zago (2016) collected data on the courts of justice system in Italy and the aim here is to make a robustness check on the findings of the aforementioned paper. For simplicity, only the data in the year 2003 are considered. Inputs in the production process are (for each court of justice) the full time equivalent number of judges and number of administrative staff. These inputs are processing cases and each year the overall number of cases processed represents the output. Two outputs are considered in order to differentiate between civil and criminal cases. Therefore a model with 2 inputs and 2 outputs is considered. From the descriptive statistics reported in Table 3 it can be seen that there is substantial variation in the quantity of resources used and the number of cases processed each year. Overall the system processed more than 2.5 million civil cases and 1.2 million criminal cases using 6,547 judges and 16,524 administrative staff.

In what follow, the base isoquant is chosen in such a way that an aggregate input and output equal to one is represented by a court of justice which uses 8 judges and 20 administrative staff to process 2,186 civil cases and 589 criminal cases. Since the input and output aggregates

	Civil Cases	Criminal Cases	Number of Judges	Admin Staff
Mean	15,507	7,538	40	100
Std Dev	24,962	10,069	60	141
Min	1,392	363	6	15
Max	216,210	76,046	492	1,232
Total	2,558,690	1,243,804	6,547	16,524

Table 3: Descriptive Statistics for 165 courts of justice in the year 2003.

are normalized to be equal to one for this observation, we can then use the input and output aggregates for all other observations to project onto this base level isoquant. Therefore we consider the adjusted dataset $\mathbf{x}_k^* = \mathbf{x}_k/X_k$ and $\mathbf{y}_k^* = \mathbf{y}_k/Y_k$. Plotting these adjusted data will show the estimate of the isoquant at all observed points for the given reference level.

In Figure 2 these data are plotted for the two outputs. Since the output isoquants are all parallel under output homotheticity, the shape of this isoquant is representative of the shape of the isoquant at all points of the technology set. In the figure the non-convex isoquant is represented in black and the convex isoquant in red. As it is clear from this picture, there are large substitution possibilities between the two outputs for a given level of resourcing. For example, a court processing about 2,200 civil cases and 800 criminal cases, could change the mix of outputs for a given level of inputs and process up to about 1,700 criminal cases by reducing the number of processed civil cases down to around 1,200. Notice how, for the convex isoquant, the substitution possibilities are almost linear.

In Figure 3 the input isoquant is plotted for the base level surface. The convex isoquant is reported in red and the non-convex isoquant in black. Since under input homotheticity all isoquants of the technology are parallel, the shape of this isoquant is representative of the input substitution possibilities. An interesting feature of this input isoquant is that the substitution possibilities between judges and administrative staff are quite limited, as one would expect. By looking at the isoquant a reduction from 8 judges to almost 7 judges requires an increase in the admin staff of about 8 units. Outside these bounds there are no substitution possibilities.

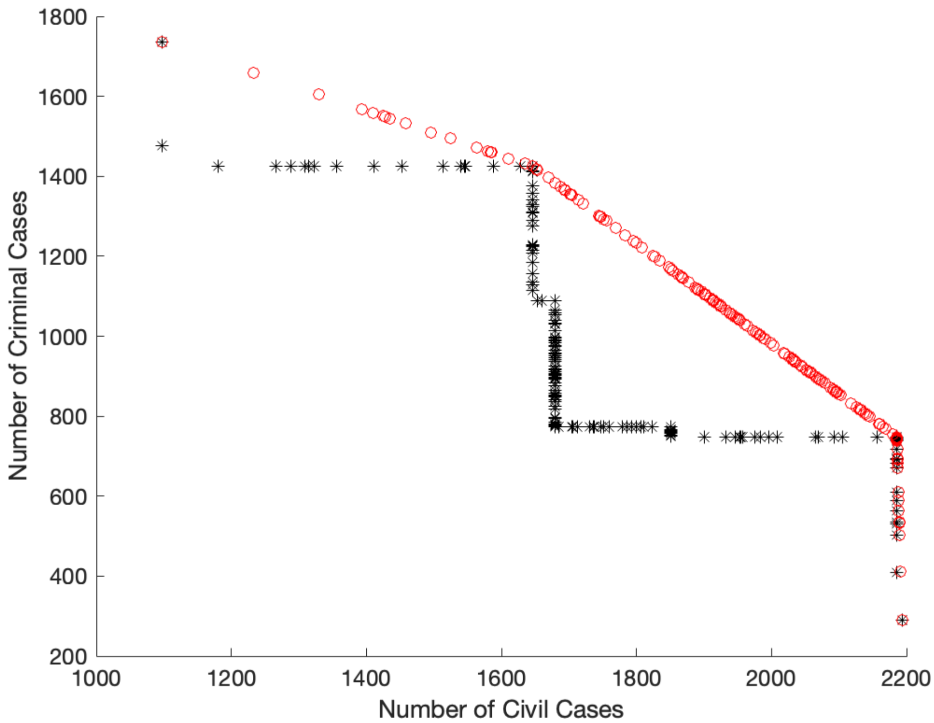


Figure 2: Output substitution possibilities under joint homotheticity.

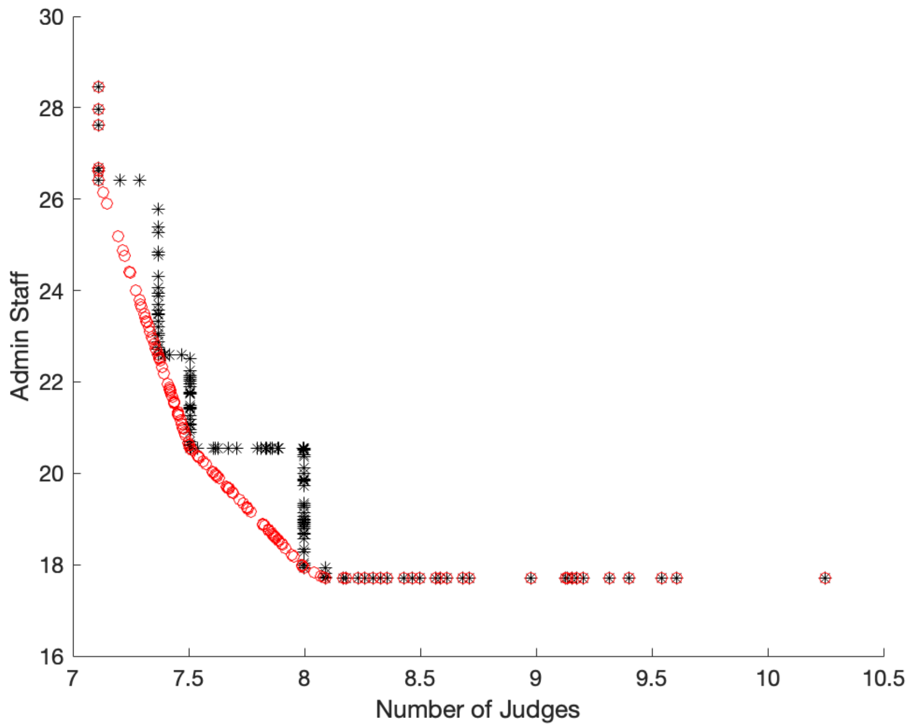


Figure 3: Input substitution possibilities under joint homotheticity.

Clearly, this isoquant is very similar to a Leontief isoquant with an optimal mix of inputs. This makes sense, since the tasks that a judge can delegate to the administrative staff are quite limited. We conclude therefore that there exists something like an "optimal" proportion of judges to administrative staff that is in the range of having between 2.25 and 3.5 administrative staff for each judge. Any ratio below or above this number will sit on the slack of both the convex and non-convex isoquant. Interestingly, the aggregate ratio is about 2.5 (considering the overall number of judges and admin staff for the whole system), well within these bounds. This suggests that there is a problem with the allocation of the inputs across courts, with too many courts showing slack inefficiency that could be used to improve the productivity of the system.

In Figure 4 a boxplot of the four sets of efficiency scores is reported. These efficiency scores are, respectively: the efficiency score under non-convexity and non-homotheticity; the score under convexity and non-homotheticity; the score under non-convexity and homotheticity; and the score under convexity and homotheticity. Notice that both homotheticity and convexity, by enlarging the technology set, will return a higher discrimination power compared to the other models.

One last piece of evidence is reported in figure 5, which represents a scatter plot of the input and output aggregates computed both under the convex and non-convex technologies for the homothetic case (non-convex in black and convex in red). Interestigly, the two sets of aggregates are not substantially different and the picture clearly shows the sort of economies of scale under which the court system operates. The black dotted line represents the CRS technology. Both the convex and the non-convex models point to the fact that there is a portion of constant returns to scale for small courts of justice, with aggregate input below a value of 15. Since a unit value of the aggregate input implies 8 judges and 20 administrative staff, the region of constant returns will be approximately below 150 judges. All courts that operate above this point will be penalized by decreasing returns to scale. This support the intuition that large

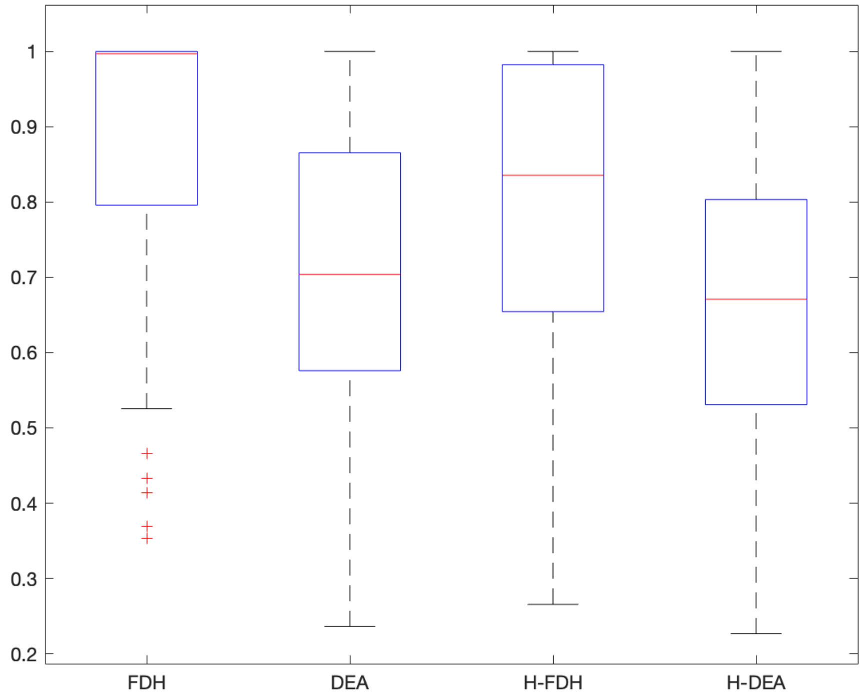


Figure 4: Comparison of output efficiency scores for the homothetic and non-homothetic case.

courts of justice should be split into smaller ones. It should also be noted that for the non-convex model, there is a strong non-convexity in the region with aggregate input between 9 and about 15, pointing to the fact that in this region there could be prevailing variable returns to scale. Therefore, according to the non-convex model, the size of the court of justice should not exceed about 72 judges.

7 Conclusion

In this study I considered estimation of a technology which exhibits joint input and output homotheticity. A method has been provided for enlarging the VRS data generated technology in order to satisfy the homotheticity assumption. The potential of the method has been illustrated by using a real dataset on courts of justice in Italy.

Since the method for imposing homotheticity is simple, numerically stable and computa-

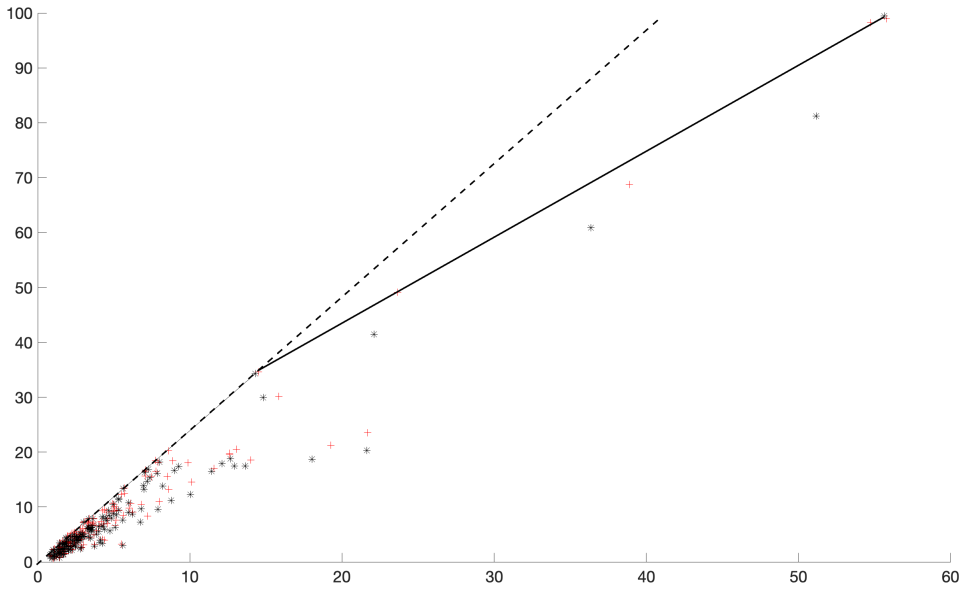


Figure 5: Frontier in the aggregate space under joint homotheticity.

tionally fast, there are various way in which this result can be used. The first possible use is to impose selective convexity on the technology. This would involve computing the aggregates using the convex CRS cone, but then, when projecting them on the aggregate space, consider a non-convex graph. This would give rise to a technology that is convex in terms of the input substitution possibilities (input isoquants), but it is non-convex in terms of scale economies (clearly the opposite can be done as well, with a non-convex isoquant, but a convex graph). In fact, one can go a step further and utilize the fact that a base isoquant exists, to impose some parametric form for the input (or output) isoquant, therefore reducing the non-parametric dimensionality of the input isoquant problem, and retaining the non-parametric non-convex structure for the graph of the technology. This would in fact be a semi-parametric production technology that satisfies all the aforementioned axioms in addition to having isoquants of a given parametric functional form.

References

Briec, W. and Kerstens, K. (2006). Input, output and graph technical efficiency measures on non-convex fdh models with various scaling laws: An integrated approach based upon implicit enumeration algorithms. *Top*, 14(1):135–166.

Hanoch, G. and Rothschild, M. (1972). Testing the assumptions of production theory: a non-parametric approach. *Journal of Political Economy*, 80(2):256–275.

Kerstens, K. and Eeckaut, P. V. (1999). Estimating returns to scale using non-parametric deterministic technologies: A new method based on goodness-of-fit. *European Journal of Operational Research*, 113(1):206–214.

O'Donnell, C. J. (2018). *Productivity and efficiency analysis*. Springer.

Olesen, O. and Ruggiero, J. (2022). The hinging hyperplanes: An alternative nonparametric representation of a production function. *European Journal of Operational Research*, 296(1):254–266.

Olesen, O. B. (2014). A homothetic reference technology in data envelopment analysis. *European Journal of Operational Research*, 233(3):759–771.

Olesen, O. B. and Ruggiero, J. (2014). Maintaining the regular ultra passum law in data envelopment analysis. *European Journal of Operational Research*, 235(3):798–809.

Peyrache, A. and Zago, A. (2016). Large courts, small justice!: The inefficiency and the optimal structure of the italian justice sector. *Omega*, 64:42–56.

Podinovski, V. (2004). On the linearisation of reference technologies for testing returns to scale in fdh models. *European Journal of Operational Research*, 152(3):800–802.

Podinovski, V. V. (2022). Variable and constant returns-to-scale production technologies with component processes. *Operations Research*, 70(2):1238–1258.

Russell, R. R. and Schworm, W. (2009). Axiomatic foundations of efficiency measurement on data-generated technologies. *Journal of Productivity Analysis*, 31(2):77–86.