

# Staring at Economic Aggregators through Information Lenses

Richard Nock

Centre d'Etude et de Recherche en Economie, Gestion, Modélisation  
et Informatique Appliquée (CEREGMIA — UAG),  
PO Box 7209, Schoelcher 97275, France.  
rnock@martinique.univ-ag.fr

Nicolas Sanz

CEREGMIA — UAG,  
PO Box 792, Cayenne 97400, France.  
Fred.Celimene@martinique.univ-ag.fr

Fred Célimène

CEREGMIA — UAG,  
PO Box 7209, Schoelcher 97275, France.  
Fred.Celimene@martinique.univ-ag.fr

Frank Nielsen

LIX — Ecole Polytechnique, Palaiseau 91128, France  
& Sony Computer Science Laboratories Inc., 3-14-13  
Higashi Gotanda, Shinagawa-Ku, 141-0022 Tokyo, Japan.  
Nielsen@acm.org

February 2, 2008

## Abstract

It is hard to exaggerate the role of economic aggregators — functions that summarize numerous and / or heterogeneous data — in economic models since the early XX<sup>th</sup> century. In many cases, as witnessed by the pioneering works of Cobb and Douglas, these functions were information quantities tailored to economic theories, i.e. they were built to fit economic phenomena. In this paper, we look at these functions from the complementary side: information. We use a recent toolbox built on top of a vast class of distortions coined by Bregman, whose application field rivals metrics' in various subfields of mathematics. This toolbox makes it possible to find the quality of an aggregator (for consumptions, prices, labor, capital, wages, etc.), from the standpoint of the information it carries. We prove a rather striking result. From the informational standpoint, well-known economic aggregators do belong to the *optimal* set. As common economic assumptions enter the analysis, this large set shrinks, and it essentially ends up *exactly fitting* either CES, or Cobb-Douglas, or both. To summarize, in the relevant economic contexts, one could not have crafted better some aggregator from the information standpoint. We also discuss global economic behaviors of optimal information aggregators in general, and present a brief panorama of the links between economic and information aggregators.

**Keywords** : Economic Aggregators, CES, Cobb-Douglas, Bregman divergences

## 1 Introduction

Since the end of the XIX<sup>th</sup> century and the birth of the “neo-classical” school, mathematics have played a growing role in economics. With the works of Léon Walras, the question of aggregation of the behavior of many individuals has risen and become central in the economic theory. In order to represent as well as possible the evolution of these aggregate variables, some mathematical functions have been proposed and become very famous in the economic literature.

One of the most famous neo-classical function is the Cobb-Douglas [7, 19]. This function is of particular interest, since it allows for perfect substitutability between the different inputs it depends on. Another well-known “linear” function was later formulated by Leontief [13], in which inputs are conversely complementary. The choice of such a function to describe the production process has very strong implications at the macroeconomic level, as illustrated by many results found by Keynesians economics in the literature on growth theory.

But beyond these different aggregate functions, one of the most recently built and well-known one is the constant elasticity of substitution (CES) function elaborated by Arrow et al. [2]. Indeed, in the Cobb-Douglas production function, the elasticity of substitution of capital for labor is fixed to unity. This implies that a one percent increase in the capital stock implies an equal one percent fall in labor inputs in order to maintain a constant production level, given the structure of relative prices. On the contrary, the CES function allows this elasticity to lie between zero and infinity, but to stay fixed at that number along and across the isoquants, whatever the quantities of inputs that are used in the production process. The main advantage exhibited by the CES function is that it encompasses the Cobb-Douglas, the Leontief and the Linear production functions, which are in fact limit and thus particular cases of it. Nevertheless, one of the reasons economists have kept on using simpler functions such as the Cobb-Douglas one is the heavy calculus to which the CES function often leads, especially at the point where models have to be closed.

In a seminal work, Douglas in [10] highlights the importance of the progresses in the field of statistical information in the genesis of his essay. Pioneering works of Cobb and Douglas [7], and Arrow *et al.* [2], underline the inductive nature of the inception of their respective functions, as the purpose was to fit as best as possible information quantities (aggregators) to observed economic phenomena. In this paper, we take a deductive route paved with a rigorous information material, to derive these fundamental quantities based on two assumptions:

- an aggregator should always be as informative as possible with respect to the data it summarizes (prices, consumptions, wages, capital, labor, etc.);
- an aggregator might be require to satisfy standard economic assumptions, relying on aggregator dualities (prices / consumptions, wages / labor, etc.), elasticities, marginal rates of substitutions, returns to scale, etc.

The starting point of our work is a class of distortions coined in the sixties by Bregman [6], in the context of convex programming. Though they were born four decades ago, it was only much later that these distortions literally spread out to other fields, including statistics, signal processing and classification [12], fields where they had to become undeniably central. It was even later that was discovered their broad applicability, with an axiomatization that makes it possible to relate them to metrics and their spawns [3]. Very roughly, *Bregman divergences* are non-negative functions that meet the same identity of indiscernibles condition as metrics, and rely on a third assumption about the existence of a particular aggregator which minimizes the total distortion to a set. This last condition, which can be rephrased as a maximum likelihood condition, makes this aggregator the *most informative* quantity about the data, and we call it a *Low Distortion Aggregator* (LDA).

In this paper, our contribution is threefold. First, we make a clear partition of economic aggregators with respect to information, as we show that some *are* LDAs (CES, Cobb-Douglas), some are limit cases of LDAs (Leontief), and some are neither (Mitscherlich-Spillman-von Thünen). Without more assumptions, the set of all LDAs is huge, yet we show that global trends of economic relevance can be easily shown for all, such as on marginal rates of substitution, and the set can be quite easily drilled down for aggregators with general behaviors, such as concavity or convexity. This, in fact, is our last contribution. Our main contribution is to show that, when we plug in various standard economic

assumptions (see above), the set of all LDAs reduces to a particular subset which *precisely* matches CES, Cobb-Douglas, or both sets. This novel advocacy for the use of these popular aggregators brings a very strong information-theoretic rationale to their “economic” existence.

The remaining of the paper is structured as follows. Section 2 presents LDAs and their main properties. In Section 3, we relate common economic aggregators to LDAs. Section 4 discusses additional properties of LDAs. A last section concludes the paper, with avenues for future research. In order not to laden the paper’s body, all proofs have been postponed to an appendix.

## 2 Low-distortion aggregators

For any strictly convex function  $\varphi : \mathbb{X} \rightarrow \mathbb{R}$  differentiable on  $\text{int}(\mathbb{X})$ , with  $\mathbb{X} \subseteq \mathbb{R}^d$  convex, the *Bregman Divergence*  $D_\varphi$  with generator  $\varphi$  is [6, 3]:

$$D_\varphi(\mathbf{x}||\mathbf{y}) \stackrel{\text{def}}{=} \varphi(\mathbf{x}) - \varphi(\mathbf{y}) - \langle \mathbf{x} - \mathbf{y}, \nabla_\varphi(\mathbf{y}) \rangle , \quad (1)$$

where  $\langle \cdot, \cdot \rangle$  denotes the inner product, and  $\nabla_\varphi \stackrel{\text{def}}{=} [\partial\varphi/\partial x_i]^\top$  is the gradient operator. In this paper, bold notations such as  $\mathbf{x}$  shall denote vector-based notations, and blackboard faces such as  $\mathbb{X}$  sets of (tuples of) real numbers or natural integers of  $\mathbb{R}$  or  $\mathbb{N}$  respectively.  $D_\varphi(\mathbf{x}||\mathbf{y})$  is the difference between the value of  $\varphi$  at  $\mathbf{x}$  and the value at  $\mathbf{x}$  of the hyperplane tangent to  $\varphi$  in  $\mathbf{y}$ . Bregman divergences encode a natural notion of distortion, as shown by Theorem 1 below. Its proof is a slight variation of Theorem 4 in [3] (see also [4]).

**Theorem 1** *Let  $F : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  be a function that satisfies the following three axioms ( $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ ):*

1. *non-negativity:*  $F(\mathbf{x}, \mathbf{y}) \geq 0$ ;
2. *identity of indiscernibles:*  $F(\mathbf{x}, \mathbf{y}) = 0$  if and only if  $\mathbf{x} = \mathbf{y}$ ;
3. *the expectation is the lowest distortion’s predictor:* for any random variable  $\mathbf{X}$  whose distribution  $\mathcal{D}$  has support  $\mathbb{R}^d$ ,

$$\arg_{\mathbf{y} \in \mathbb{R}^d} \min \mathbf{E}_{\mathcal{D}} F(\mathbf{X}, \mathbf{y}) = \mathbf{E}_{\mathcal{D}} \mathbf{X} \stackrel{\text{def}}{=} \mu_{\mathbf{X}} , \quad (2)$$

where  $\mathbf{E}_{\mathcal{D}}$  denotes the mathematical expectation.

Then  $F(\mathbf{x}, \mathbf{y}) = D_\varphi(\mathbf{x}||\mathbf{y})$  for some strictly convex and differentiable  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ .

(proof: see the Appendix) It is easy to check that any Bregman divergence satisfies [1], [2] and [3] [4], and so Theorem 1 provides a complete characterization of Bregman divergences, in the same way as conditions [1] and [2], completed with symmetry and subadditivity, would axiomatize a metric. This positions Bregman divergences with respect to numerous metric-related notions, and gives the importance of their main difference, eq. (2). Eq. (2) is fundamental because it says that the (arithmetic) expectation is the lowest distortion parameter for a population, regardless of the distortion. Actually, eq. (2) says much more: the expectation is *maximum likelihood estimator* of data for a large set of distributions called the *exponential families*. These families contain some of the most popular distributions, such as Bernoulli, multinomial, beta, gamma, normal, Rayleigh, Laplacian, Poisson [4, 15]. A remarkable property is that any member satisfies the following identity [4]:

$$\log \mathbf{Pr}[\mathbf{x}|\boldsymbol{\theta}, \varphi] = -D_\varphi(\mathbf{x}||\boldsymbol{\mu}_\boldsymbol{\theta}) + \log b_\varphi(\mathbf{x}) . \quad (3)$$

$\text{dom}(\varphi)$	$\phi(x)$	$D_\varphi(\mathbf{x}  \mathbf{y})$	Divergence name
$\mathbb{R}^d$	$x^2$	$\sum_{i=1}^d (x_i - y_i)^2$	Squared Euclidean norm
$\mathbb{R}_+^d$	$x \log x - x$	$\sum_{i=1}^d x_i \log \frac{x_i}{y_i} - x_i + y_i$	Kullback-Leibler div.
$\mathbb{P}_d$	id.	$\sum_{i=1}^d x_i \log \frac{x_i}{y_i}$	Entropy
$\mathbb{R}_+^d$	$-\log x$	$\sum_{i=1}^d \frac{x_i}{y_i} - \log \frac{x_i}{y_i} - 1$	Itakura-Saito div.
$[0, 1]$	$x \log x + (1-x) \log(1-x)$	$x \log \frac{x}{y} + (1-x) \log \frac{1-x}{1-y}$	Logistic loss

Table 1: Correspondence between various generators and their Bregman divergences.  $\mathbb{P}_d$  is the  $d$ -dimensional probability simplex. The generator of the Bregman divergence is defined by  $\varphi(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{i=1}^d \phi(x_i)$  (see text).

$\boldsymbol{\theta}$  defines the so-called natural parameters of the distribution, and  $b_\varphi(\cdot)$  is a normalization function. It follows from (3) and (2) that the maximum likelihood estimator of data is the expectation parameter  $\boldsymbol{\mu}_\theta$ .

Some Bregman divergences have become cornerstones of various fields of mathematics and computer science, as shown in Table 1. All of them are *separable* Bregman divergences [8], as they can be characterized using a strictly convex function  $\phi : \mathbb{I} \subseteq \mathbb{R} \rightarrow \mathbb{R}$ , the generator of the Bregman divergence being just:

$$\varphi(\mathbf{x}) \stackrel{\text{def}}{=} \sum_i \phi(x_i) . \quad (4)$$

It might seem that (2) unveils a strong asymmetry between the two parameters of a Bregman divergence, all the more as that Bregman divergences are not symmetric in almost all cases [15]. This distinction becomes more superficial — but crucial for our purpose — as *Legendre duality* enters the analysis. Any Bregman divergence is indeed equal to a Bregman divergence over *swapped* parameters in the generator’s gradient space. To make it formal, the generator  $\varphi$  of a Bregman divergence admits a convex conjugate  $\varphi^* : \mathbb{R}^d \rightarrow \mathbb{R}$  given by [17]:

$$\varphi^*(\mathbf{y}) \stackrel{\text{def}}{=} \sup_{\mathbf{x} \in \mathbb{X}} \{ \langle \mathbf{x}, \mathbf{y} \rangle - \varphi(\mathbf{x}) \} \quad (5)$$

$$= \langle \mathbf{y}, \nabla_\varphi^{-1}(\mathbf{y}) \rangle - \varphi(\nabla_\varphi^{-1}(\mathbf{y})) , \quad (6)$$

where  $\nabla_\varphi^{-1}$ , the inverse gradient, is well-defined because of the strict convexity of  $\varphi$ . The following Theorem, whose proof follows from plugging (6) in (1), states the dual symmetry of Bregman divergences.

**Theorem 2**  $D_\varphi(\mathbf{x}||\mathbf{y}) = D_{\varphi^*}(\nabla_\varphi(\mathbf{y})||\nabla_\varphi(\mathbf{x}))$ .

It follows from (2) and the strict convexity of  $\varphi$  that the minimizer of the expected dual divergence  $D_{\varphi^*}(\cdot||\cdot)$  can be expressed in  $\text{int}(\text{dom}(\varphi))$  as:

$$\boldsymbol{\mu}_\varphi \stackrel{\text{def}}{=} \nabla_\varphi^{-1}(\mathbf{E}_\mathcal{D} \nabla_\varphi(\mathbf{X})) . \quad (7)$$

The set spanned by (7), which includes the arithmetic average (take  $\phi \stackrel{\text{def}}{=} x^2/2$  in (4)), is close to the set of  $f$ -means [11], a set whose studies date back to the early thirties, by Kolmogorov and Nagumo.

To summarize the conceptual justifications for the use of *aggregators* having shape (7), three main motivations could justify their use: first, they are all optimal distortion estimators — and the only

ones to be optimal – in the sense of Theorem 1; second, they encode maximum likelihood estimators for a majority of popular distributions; third, they encode geodesic-like curves in the geometry of the information space [15]. For all these reasons, they can be considered the best information aggregators for the data they summarize (data which could be prices, consumptions, labors, wages, etc. in the economic world).

Hereafter, we consider averages (7) with finite support of size  $m > 0$ , and replace (2) by the more general search for  $\arg_{\mathbf{y} \in \mathbb{R}^d} \min \sum_{i=1}^m \gamma_i F(\mathbf{x}_i, \mathbf{y})$ , with  $\gamma_i > 0, \forall i = 1, 2, \dots, m$ . A rapid glimpse at (2) reveals that the solution is  $\sum_{i=1}^m \gamma_i \mathbf{x}_i$ , and so the extension of (7) to the minimizer of a general weighted sum of Bregman divergences now takes the more general form:

$$\boldsymbol{\mu}_\varphi \stackrel{\text{def}}{=} \Gamma \nabla_\varphi^{-1} \left( \frac{1}{\Gamma} \sum_{i=1}^m \gamma_i \nabla_\varphi(\mathbf{x}_i) \right), \quad (8)$$

with  $\Gamma \stackrel{\text{def}}{=} \sum_{i=1}^m \gamma_i$ . Because of (2), any  $\boldsymbol{\mu}_\varphi$  as in (8) is called a low-distortion aggregator (LDA). For economic and mathematical reasons, averages having the form (8) with a concave or convex regime are particularly interesting. The following Theorem allows to catch the picture of where concavity and convexity lie: the symmetric dual average of some average (8) enjoys the symmetric regime. If one is concave, the other is convex and *vice versa*.

**Theorem 3**  $\mu_\varphi$  is concave if and only if  $\mu_{\varphi^*}$  is convex.

(proof: see the Appendix). To finish up with information, we state the last result that shall be useful in the sequel.

**Theorem 4** Let  $\mu_\varphi$  a concave (resp. convex) average that follows (7). Then it is upperbounded (resp. lowerbounded) by the sum:  $\mathbf{s} = \sum_{i=1}^m \gamma_i \mathbf{x}_i$ . Furthermore,  $\nabla_\varphi$  is concave (resp. convex), and  $\nabla_\varphi^{-1}$  is convex (resp. concave).

(proof: see the Appendix).

### 3 Economic Aggregators

Because a LDA  $\mu_\varphi$  does not change by adding a constant term to its generator  $\varphi$ , it should be kept in mind that generators shall be given up to any such constant. Furthermore, our analysis takes place for separable generators, that meet (4). This eases readability while encompassing most economic settings. For such reasons, it is also convenient to assume that  $\text{dom}\phi \subseteq \mathbb{R}_+$ , and introduce the following notation for any relevant  $k \in \mathbb{N}_*$ :

$$\phi^{[k]}(x) \stackrel{\text{def}}{=} \frac{d^k \phi(x)}{dx^k}. \quad (9)$$

#### 3.1 Optimality of Economic Aggregators

Let  $x_\star$  denote an aggregator for values  $x_1, x_2, \dots, x_m$  ( $m \in \mathbb{N}_*$ ). One of the most common economic aggregators is the CES function [2]:

$$x_\star \stackrel{\text{def}}{=} \left( \sum_{i=1}^m \beta_i x_i^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}. \quad (10)$$

The formulation in  $\sigma$  is not the simplest but it is intentional, as it depicts the constant elasticity of substitution inside values aggregated [5]. Here,  $\beta_i > 0$  is the weight of aggregated value  $x_i$ . Further

constraints of economic relevance are generally imposed on  $\sigma$  depending on the setting in which (10) is applied [5]; in order to remain as general as possible, we consider the unrestricted setting for which  $\sigma \in \mathbb{R}_* \setminus \{1\}$ . We now show that a CES is a LDA.

**Lemma 1** *Any CES  $x_\star$  as defined in (10) is a LDA for the generator*

$$\phi_{\text{CES}}(x) \stackrel{\text{def}}{=} ax^{2-\frac{1}{\sigma}} , \quad (11)$$

with  $a \in \mathbb{R}_*$  any constant for which (11) is convex, and  $\gamma_i \stackrel{\text{def}}{=} \beta_i B^{\frac{1}{\sigma-1}}, \forall i = 1, 2, \dots, m$ . Here,  $B \stackrel{\text{def}}{=} \sum_{i=1}^m \beta_i$ .

(proof: see the Appendix). Aggregators are sometimes tied up via important economic equalities. One example relates prices and consumptions. Let  $m$  denote the number of goods, and the consumption function of good  $i$  is noted  $c_i$ . The price of good  $i$  is  $p_i$ . Two aggregators for consumptions and prices, respectively  $c_\star$  and  $p_\star$  are devised so as to satisfy:

$$\sum_{i=1}^m c_i p_i = p_\star c_\star . \quad (12)$$

Further economic assumptions can be made, such as the concavity of  $c_\star$ , which indicates the preference for diversity [9]. The popular choice for  $c_\star$  is a CES function (10) [2]. Notice that the weights in the LDA ( $\gamma_i$ ) are different from the weights in the CES ( $\beta_i$ ). Modulo a simple normalization of the CES, they remain equal. If we multiply  $c_\star$  by  $B^{1/(1-\sigma)}$ , the normalized CES obtained is such that  $\gamma_i = \beta_i$ . Furthermore, this normalization, for which  $B^{1/(1-\sigma)} = m^{1/(1-\sigma)}$  when all  $\beta_i = 1$ , is one which turns out to play a key role in economic models [5].

The price aggregator,  $p_\star$ , can be found by inspecting (12) after remarking that partial derivatives on the left and right-hand side must also coincide. After a standard derivation using (10) for  $c_\star$ , we obtain that the price index has the form:

$$p_\star = \left( \sum_{i=1}^m \beta_i^\sigma p_i^{1-\sigma} \right)^{\frac{1}{1-\sigma}} . \quad (13)$$

$p_\star$  has also the general CES form of (10); for completeness, we characterize below its LDA (proof similar to Lemma 1).

**Lemma 2** *Let  $\delta_i \stackrel{\text{def}}{=} \beta_i^\sigma$ , and  $\Delta \stackrel{\text{def}}{=} \sum_{i=1}^m \delta_i$ . The price index in (13) is a LDA for the generator:*

$$\phi_p(x) \stackrel{\text{def}}{=} bx^{2-\sigma} , \quad (14)$$

with  $b \in \mathbb{R}_*$  any constant for which (14) is convex, and  $\gamma_i \stackrel{\text{def}}{=} \delta_i \Delta^{\frac{\sigma}{1-\sigma}}, \forall i = 1, 2, \dots, m$ .

Modulo the normalization of the CES for  $c_\star$ , and the choice  $\beta_i = 1$ , (13) would return to the conventional choice in which  $\beta_i^\sigma \rightarrow 1/m$ . It is quite a remarkable fact that  $p_\star$  and  $c_\star$  are LDA under the sole assumptions of (12) and  $c_\star$  is a CES. Such a property also holds for labor and wages. Suppose we have  $n$  consumer-workers, each of which selling a particular labor type; let  $w_j$  be the wage for labor-type  $j$  and  $n_j$  the demand for labor-type  $j$ , for  $j = 1, 2, \dots, n$ . Then there exists an aggregate labor-demand index  $n_\star$ , and a wage index  $w_\star$ , such that [5]:

$$\sum_{j=1}^n w_j n_j = w_\star n_\star . \quad (15)$$

The CES form for  $w_\star$  [5] implies both the LDA property for  $w_\star$  and  $n_\star$  (Lemmata 1 and 2). To summarize, popular aggregators for consumptions, prices, labor and wages are all LDAs, which means that they are all optimal from the information theory standpoint. Before drilling down further into the properties that yield relationships like (12) or (15), let us give a brief panorama of which Bregman divergences are involved so far.

The Bregman divergence of a CES (10) is:

$$D_{\phi_{\text{CES}}}(\mathbf{x}||\mathbf{z}) = a \sum_{i=1}^m \left\{ x_i^{2-\frac{1}{\sigma}} - \left(2 - \frac{1}{\sigma}\right) x_i z_i^{1-\frac{1}{\sigma}} + \left(1 - \frac{1}{\sigma}\right) z_i^{2-\frac{1}{\sigma}} \right\}. \quad (16)$$

Since any CES is a LDA, it follows that Cobb-Douglas and Leontief functions are limit LDAs, respectively when  $\sigma \rightarrow 1$  and  $\sigma \rightarrow 0^+$ . While Leontief function,  $x_\star \stackrel{\text{def}}{=} \min_i \{\beta_i x_i\}$ , does not admit a generator (it is not differentiable), Cobb-Douglas,

$$x_\star \stackrel{\text{def}}{=} \prod_{i=1}^m x_i^{\beta_i}, \quad (17)$$

admits one, which is:

$$\begin{aligned} \phi_{\text{CD}}(x) &\stackrel{\text{def}}{=} b(x \log x - x) \\ &= b\phi_{\text{KL}}(x) \end{aligned} \quad (18)$$

(see Table 1;  $b \in \mathbb{R}_{+,*}$  is any constant). If we look at the price index in (13), we get the following result.

**Lemma 3** Fix  $b = 1/((2 - \sigma)(1 - \sigma))$  in (14), assuming  $\sigma \neq 1$ , and let  $\phi_{\text{IS}} \stackrel{\text{def}}{=} -\log x$  (see Table 1). Then

$$\lim_{\sigma \rightarrow 2} D_{\varphi_p}(\mathbf{x}||\mathbf{z}) = D_{\varphi_{\text{IS}}}(\mathbf{x}||\mathbf{z}). \quad (19)$$

This result is easily proven once we remark that  $x^k \approx 1 + k \log x + o(k)$ . The right-hand side of (19) is Itakura-Saito divergence (Table 1). Together with the fact that the limit divergence for  $c_\star$  is Kullback-Leibler divergence when  $\sigma \rightarrow 1$ , we get the generators for two popular divergences of signal processing and statistics [15]. A well-known similar result holds for a particular subset of Bregman divergences, Amari  $\alpha$ -divergences, for which [1]:

$$\phi_{\text{A}}(x) \stackrel{\text{def}}{=} 4 \left( x - x^{\frac{1+\alpha}{2}} \right) / (1 - \alpha^2), \alpha \in [-1, 1]. \quad (20)$$

Taking limits of the generator when  $\alpha$  reaches the interval bounds yields Itakura-Saito and Kullback-Leibler divergences.

### 3.2 Completeness of Economic Aggregators

In this section, we consider some relevant economic assumptions about aggregators, and show that any LDA that would meet such assumptions would necessarily belong to a particular subclass of LDAs. This subclass is called “complete” for the assumption at hand.

The first assumption we consider is about any two *dual* aggregators  $x_\star$  (for  $x_1, x_2, \dots, x_m$ ) and  $z_\star$  (for  $z_1, z_2, \dots, z_m$ ) that would meet the following abstraction of (12) and (15):

$$\sum_{i=1}^m x_i z_i = x_\star z_\star. \quad (21)$$

We show that CES turns out to be complete for dual aggregators, as the LDA assumption for any of the two implies that *both* are CES. We state it more formally below.

**Theorem 5** *Suppose that at least one of  $x_\star$  and  $z_\star$  that satisfies (21) is a LDA. Then both  $x_\star$  and  $z_\star$  are CES. Furthermore, they are linked through the identity  $\phi_z^{[2]} = d \left( \phi_x^{[2]} \right)^{-1}$  for some  $d \in \mathbb{R}_\star$ .*

(proof: see the Appendix). CES turns out to be complete from another standpoint: elasticities. Consider some LDA  $x_\star$ ; its elasticity with respect to  $x_i$  ( $i = 1, 2, \dots, m$ ) is defined as:

$$E_{x_\star}^{x_i} \stackrel{\text{def}}{=} \left( \frac{dx_\star}{x_\star} \right) / \left( \frac{dx_i}{x_i} \right) . \quad (22)$$

Consider the economic assumption that all elasticities sum to one. We show that CES is complete for this assumption.

**Theorem 6** *Let  $x_\star(x_1, x_2, \dots, x_m)$  be any LDA. Then  $\sum_{i=1}^m E_{x_\star}^{x_i} = 1$  if and only if  $x_\star$  is a CES.*

(proof: see the Appendix). We now switch to another important economic quantity, the substitution elasticity of  $x_i$  for  $x_j$  in  $x_\star$ ,  $E_{x_\star}^{x_i \rightarrow x_j}$ , defined by:

$$E_{x_\star}^{x_i \rightarrow x_j} \stackrel{\text{def}}{=} \left( \frac{d(x_j/x_i)}{x_j/x_i} \right) / \left( \frac{dS_{x_\star}^{x_i \rightarrow x_j}}{S_{x_\star}^{x_i \rightarrow x_j}} \right) , \quad (23)$$

where

$$S_{x_\star}^{x_i \rightarrow x_j} \stackrel{\text{def}}{=} \left( \frac{\partial x_\star}{\partial x_i} \right) / \left( \frac{\partial x_\star}{\partial x_j} \right) \quad (24)$$

is the marginal rate of substitution of  $x_i$  for  $x_j$ . Another economic assumption commonly encountered is the fact that  $E_{x_\star}^{x_i \rightarrow x_j}$  is assumed to be unit. We show that the complete LDA subclass for this assumption is, this time, Cobb-Douglas.

**Theorem 7** *Let  $x_\star(x_1, x_2, \dots, x_m)$  be any LDA. Then, there exists indices  $1 \leq i, j \leq m$  such that  $E_{x_\star}^{x_i \rightarrow x_j} = 1$  if and only if  $x_\star$  is a Cobb-Douglas.*

(proof: see the Appendix). It is interesting to notice that the LDA assumption competes with the homogeneity assumptions about  $x_\star$  that are required to come up with the same result (i.e. without making the LDA assumption). The fact that we are able to alleviate the economic setting (homogeneity ties up  $x_\star$  with assumptions on returns to scale) while ending up with the same aggregator makes information a very valuable companion to introduce the true nature of popular economic aggregators. One question which remains is however what would imply the homogeneity assumption *alone* in a LDA setting. We define  $x_\star$  to be homogeneous of degree  $a \in \mathbb{R}_\star$  if and only if:

$$x_\star(\lambda x_1, \lambda x_2, \dots, \lambda x_m) = \lambda^a x_\star(x_1, x_2, \dots, x_m) , \quad (25)$$

for every  $\lambda \in \mathbb{R}_+$ . We show that the complete LDA subclass for this assumption varies depending on the values of  $a$ . Without losing too much generality, the Theorem assumes that  $\phi_x^{[2]}$  is differentiable.

**Theorem 8** *Let  $x_\star(x_1, x_2, \dots, x_m)$  be any LDA, and  $a \in \mathbb{R}_\star$ . Then:*

- $x_\star$  is homogeneous of degree  $a \neq 1$  if and only if it is a Cobb-Douglas;
- $x_\star$  is homogeneous of degree  $a = 1$  if and only if it is a Cobb-Douglas or a CES.

(proof: see the Appendix).



	Optimality (LDA)	Completeness				
		Th. 5	Th. 6	Th. 7	Th. 8 ( $a \neq 1$ )	Th. 8 ( $a = 1$ )
CES	Y	Y	Y	N	N	Y
Cobb-Douglas	Y	N	N	Y	Y	Y
Leontief	L	L	L	N	N	L
MST	N	N	N	N	N	N

Table 2: Summary of our results on four families of aggregators: CES, Cobb-Douglas, Leontief and MST, with respect to the assumptions made in Theorems 5, 6, 7 and 8 (see text for details).

## 4 Discussion

**Families of economic aggregators** Table 2 summarizes the results obtained on three families of aggregators: CES, Cobb-Douglas and Leontief. For each of them we give the indication of whether they are LDAs (Y/N), whether they can be in the limit (L), and whether they become complete for the assumptions made in Theorems 5, 6, 7 and 8 (Y / N / L). Remark that the Table makes a clear distinction between all these three families of aggregators. There exists various other aggregators in economic works; for obvious space reasons, we have chosen to focus on the most popular, and it turns out that all have strong relationships with LDAs, either directly, or at the limit. In order to cover the possible relationships between aggregators and LDAs, let us take a last example, of a general class of aggregators that we call Mitscherlich-Spillman-von Thünen (MST) aggregators [14, 18, 19], a family in which the global form of aggregator  $x_\star$  reduces directly or after a variable change to ( $\theta \in \{-1, +1\}$ ):

$$x_\star \stackrel{\text{def}}{=} \prod_{i=1}^m (1 - \exp(\theta \gamma_i x_i)) . \quad (26)$$

Such aggregators date back to the XIX<sup>th</sup> century, and so they have preceded those we have been focusing on so far. What we can show is that, contrasting with their successors, MST aggregators are *not* LDAs.

**Lemma 4** *MST aggregators are not LDAs.*

(proof: see the Appendix).

**Aggregators and economic constraints** Modulo changes of variables, Theorems (5) - (8) could be alleviated from the constraint of the LDA choice under their respective economic assumptions. Consider for example (21), in which we would like to plug any LDA. To be concrete, let us stick to prices and consumptions in (12). Consider the generator  $\phi_{\tilde{c}}$  of some strictly concave LDA  $\tilde{c}_\star$  that aggregates its consumptions  $\tilde{c}_i$  for  $i = 1, 2, \dots, m$ . Let us say that strict concavity is chosen because usual consumption indexes are concave, to indicate the consumer’s preference for diversity [9]. Consider the change of variables that involves some (concave) CES:

$$c_i \stackrel{\text{def}}{=} \left( \phi_{\text{CES}}^{[1]} \right)^{-1} \left( \phi_{\tilde{c}}^{[1]}(\tilde{c}_i) \right) . \quad (27)$$

If we consider that consumptions  $c_i$  are the actual observed consumptions (the  $\tilde{c}_i$ ’s being “hidden”, non-observed consumptions), the aggregator  $c_\star$  for those lifted consumptions is a CES which may be

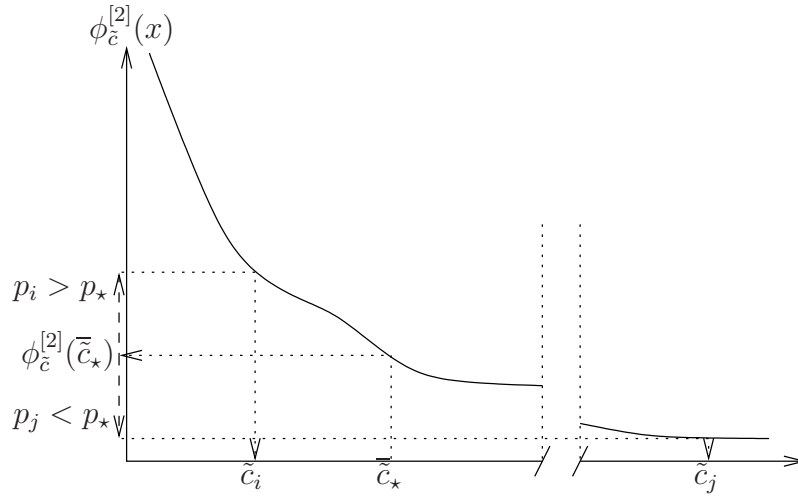


Figure 1: Prices vs consumptions for a concave consumption LDA in (30): regardless of the LDA, for any good  $i$ , if its price is larger (resp. smaller) than the price index, its consumption cannot be larger (resp. smaller) than the normalized consumption index (see text for details).

plugged in (12). Reconstructing  $\tilde{c}_\star$  is immediate as we have:

$$\tilde{c}_\star = \left( \phi_{\tilde{c}}^{[1]} \right)^{-1} \left( \phi_{\text{CES}}^{[1]}(c_\star) \right) . \quad (28)$$

In order not to laden the discussion, let us consider that all weights  $\beta_i = 1/m$ , and that we keep the CES form for  $p_\star$  ( $p_i = \tilde{p}_i$ ). What interests us here is not exactly the consumption index ( $c_\star$ ), but the *normalized* index,  $\bar{c}_\star \stackrel{\text{def}}{=} c_\star/m$ , which is really homogeneous to the consumption of a single good. We also define the same index  $\tilde{c}_\star$  for  $\tilde{c}_\star$ , and assume the same relationship as (28) for these two indices. Differentiating (12) in  $c_i$  yields:

$$\frac{p_i}{p_\star} = \frac{\phi_{\text{CES}}^{[2]}(c_i)}{\phi_{\text{CES}}^{[2]}(c_\star)} , \quad (29)$$

from which we obtain using (27) and (28):

$$\frac{p_i}{p_\star} = \frac{\phi_{\tilde{c}}^{[2]}(\tilde{c}_i)}{\phi_{\tilde{c}}^{[2]}(\tilde{c}_\star)} . \quad (30)$$

Eq. (30) is interesting because it displays a remarkably stable behavior that holds for *any* concave LDA. Because of Theorem 4 and the convexity of  $\phi$ ,  $\phi_{\tilde{c}}^{[2]}$  is monotonic decreasing and strictly positive. Thus, it converges towards some non negative value. Figure 1 gives an overview of how prices and consumptions always behave. There are two conclusions to draw from the figure. The first is a sanity check, as larger prices mean lower consumptions, a conclusion that follows from dividing (30) for distinct goods  $i$  and  $j$ . Indeed, if a good  $i$  has price  $p_i > p_j$ , then we shall have  $\tilde{c}_i \leq \tilde{c}_j$  (notice, from (27), that it is equivalent to saying  $c_i \leq c_j$ ). We also remark that prices that are larger (resp. smaller) than the price index mean consumptions that are smaller (resp. larger) than the normalized consumption index (see Figure 1). The second is the dampening effect of prices on consumptions: a small difference on prices may incur a very large difference on consumptions if those prices are already small, and it can make almost no difference on consumptions if prices are high.

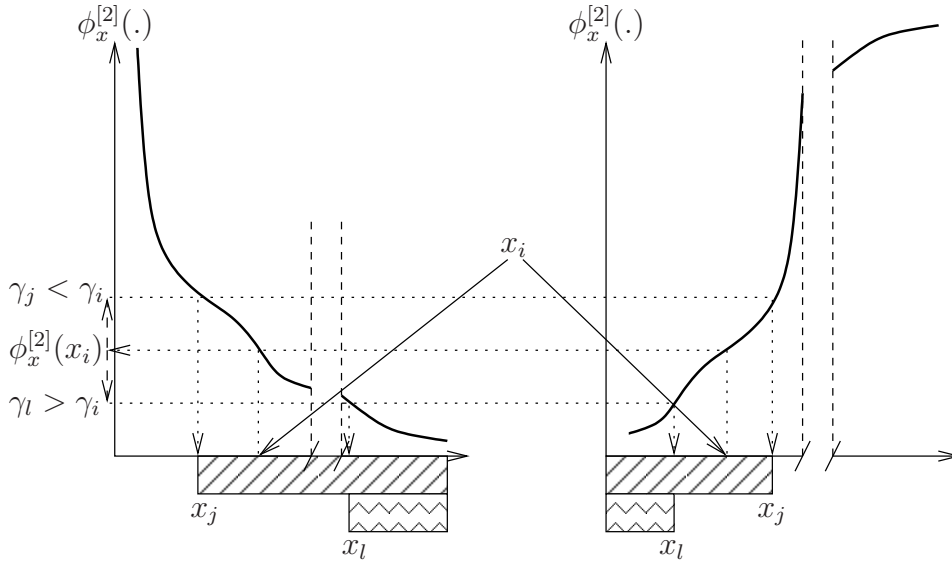


Figure 2: Intervals for which the marginal rate of substitution of  $x_i$  for  $x_j$  (or  $x_l$ ) exceeds unit when  $x_*$  is a concave (left) or convex (right) LDA.

**Global behaviors of LDAs as economic aggregators** Even without a change of variables, LDAs sometimes display economic regimes with extremely close behaviors, as witnessed by the marginal rate of substitution of  $x_i$  for  $x_j$ . Indeed, whenever  $x_*$  is a LDA, we have:

$$s_{x_*}^{x_i \rightarrow x_j} = \frac{\gamma_i \phi_x^{[2]}(x_i)}{\gamma_j \phi_x^{[2]}(x_j)} . \quad (31)$$

Figure 2 displays the general behaviors of  $s_{x_*}^{x_i \rightarrow x_j}$  as a function of the concavity or convexity of  $x_*$ . In the convex case, Theorem 4 and the convexity of  $\phi$  bring that  $\phi_x^{[2]}$  is monotonic increasing and strictly positive, hence the schema depicted in Figure 2 (right). The dashed rectangles depict the intervals for which this marginal rate of substitution would be greater than 1, that is, locations where we would be willing to trade more than one unit of  $x$  to obtain one unit of  $x_i$ . The behavior is remarkably linked with the global regime of  $x_*$ : when it is concave, Figure 2 (left) clearly displays a preference for diversity, while when it is convex, Figure 2 (right) shows the symmetric trend, an aversion for diversity.

**Aggregators of aggregators and economic programs** LDAs may incorporate heterogeneous quantities and even LDAs as well, as it is common for economic aggregators to integrate other economic aggregators. Examples show how the whole aggregate may behave, and how global regimes underlined above for “baseline” aggregators also emerge in a simple manner for whole aggregates as well. Consider the determination of the global consumption index  $c_*$  and money expenses  $m$  (hereafter,  $m$  does not refer anymore to the number of aggregated values) of a consumer, based on his/her whole budget  $r$  and price indices  $p_*$  (we do not make any assumption on the form of  $c_*$  and  $p_*$ ). The consumer solves the maximization of a utility aggregator  $u_*$ :

$$\max_{c_*, m} u_* \quad \text{s.t.} \quad p_* c_* + m = r . \quad (32)$$

Let us investigate the general solution of (32), under the sole assumption that  $u_\star$  is some concave LDA that mixes consumption and money via some Bernoulli distribution  $B(\gamma)$  for coefficients  $\gamma_i, i = 1, 2$  which leverages the importance of consumption and money in  $u_\star$ :

$$u_\star = \nabla_{\phi_u}^{-1}(\gamma \nabla_{\phi_u}(c_\star) + (1 - \gamma) \nabla_{\phi_u}(m/p_\star)) . \quad (33)$$

The following Theorem states the fundamental relationships that may be used to determine  $c_\star$  and  $m$ , for any concave LDA.

**Theorem 9** *The optimal values for  $c_\star$  and  $m$  in (32) satisfy:*

$$\phi_u^{[2]}(c_\star) = \frac{1 - \gamma}{\gamma} \phi_u^{[2]}\left(\frac{r}{p_\star} - c_\star\right) , \quad (34)$$

$$\phi_u^{[2]}\left(\frac{m}{p_\star}\right) = \frac{\gamma}{1 - \gamma} \phi_u^{[2]}\left(\frac{r}{p_\star} - \frac{m}{p_\star}\right) . \quad (35)$$

(proof: see the Appendix). Because  $\phi_u^{[2]}$  is monotonous decreasing and strictly positive, solving (34) and (35) can be done via a simple dichotomic search in the general case where  $\phi^{[2]}(x)$  would be strictly monotonous (when strict monotonicity is not ensured, we may logically end up with an interval of values). The proof of Theorem 9 reveals an interesting relationship between consumption and money, namely:

$$\phi_u^{[2]}(c_\star) = \frac{1 - \gamma}{\gamma} \phi_u^{[2]}\left(\frac{m}{p_\star}\right) . \quad (36)$$

Figure 3 displays this relationship, and more precisely where  $m/p_\star$  is located with respect to  $c_\star$ , depending on  $\gamma$ . Remark that when  $\gamma > 1/2$ , which marks the predominance of consumption over money in the consumers' utility function  $u_\star$ , Figure 3 shows that  $c_\star$  indeed represents more than money in the whole budget, as we recall that  $c_\star + (m/p_\star) = r/p_\star$ . The symmetric situation holds when  $\gamma < 1/2$ .

## 5 Conclusion

The aim of this paper was to demonstrate that the choice of various economic aggregators, that have mainly been originally built on empirical economic grounds, meet optimality from the information standpoint, and most notably, remain the *only* optimal aggregators as various standard economic assumptions are considered. In these settings, they are the only ways one could summarize at best numerous individual variables in aggregate indexes, designed to represent their collective economic behavior at the global level. We have also displayed the fact that information aggregators meet consistent economic behaviors for fairly general settings, and that these behaviors may be extracted using simple derivations.

Our results might have applications, and implications, on various economic areas whose theoretical and empirical results depend on the use of functions embodied in the LDAs studied in this paper. This includes, for instance, theories such as imperfect competition, international trade and growth theory. LDAs might make it possible to transfer, in the economic frameworks, three main advantages that Bregman divergences in general have extensively brought to statistics, classification and even geometry [4, 15, 16]. First, the analytical expression of a LDA can be extremely complex, yet the abstraction of the general form (8), and its properties (Theorems 3, 4), makes it possible for fairly complex behaviors to be derived in a simple manner. Second, it has been shown in the aforementioned fields that algorithms for solving different problems on a single divergence could be generalized to solving the same problems on any divergence. Last, but not least, LDAs encode such a large variety of functions that they could make it easier to find tight fits to economic data.

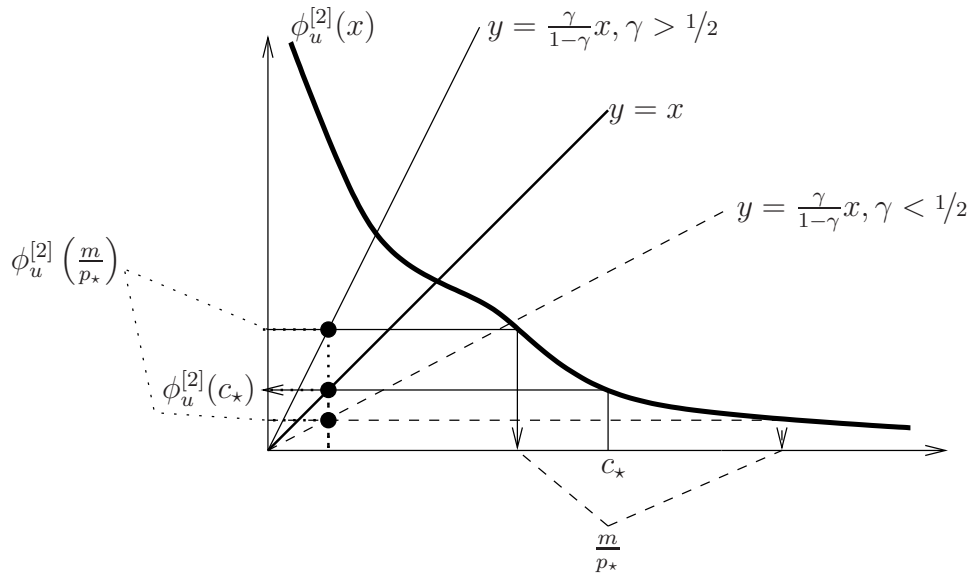


Figure 3: Depiction of the relationships between consumption and money for the consumer’s program (32) (when  $\gamma > 1/2$  and  $\gamma < 1/2$ ; see text for details).

## Acknowledgments

The authors thank Pierre Cahuc for having read an earlier draft of this paper. R. Nock and N. Sanz gratefully acknowledge support from the State Secretary of Ultraperipheral Regions through grant 07MAR15 “Coordination Failures and Distortions in Economic Models”. F. Nielsen and R. Nock are supported by National Research Agency (ANR) Blanc project ANR-07-BLAN-0328-01 “Computational Information Geometry and Applications”.

## References

- [1] S.-I. Amari and H. Nagaoka. *Information Geometry*. Cambridge University Press, 1999.
- [2] K.-J. Arrow, H.-B. Chenery, B.-S. Minhas, and R.-M. Solow. Capital-labor substitution and economic efficiency. *Review of Economics and Statistics*, 43:225–250, 1961.
- [3] A. Banerjee, X. Guo, and H. Wang. On the optimality of conditional expectation as a bregman predictor. *IEEE Trans. on Information Theory*, 51:2664–2669, 2005.
- [4] A. Banerjee, S. Merugu, I. Dhillon, and J. Ghosh. Clustering with bregman divergences. *Journal of Machine Learning Research*, 6:1705–1749, 2005.
- [5] O.-J. Blanchard and N. Kiyotaki. Monopolistic Competition and the Effect of Aggregate Demand. *American Economic Review*, 77:647–666, 1987.
- [6] L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comp. Math. and Math. Phys.*, 7:200–217, 1967.
- [7] C. Cobb and P.-H. Douglas. A theory of production. *American Economic Review*, 18:139–165, 1928.

- [8] I. Dhillon and S. Sra. Generalized non-negative matrix approximations with Bregman divergences. In *Advances in Neural Information Processing Systems 18*, 2005.
- [9] A.-K. Dixit and J.-E. Stiglitz. Monopolistic competition and optimum product diversity. *American Economic Review*, 67:297–308, 1977.
- [10] P.-H. Douglas. *The Theory of Wages*. Macmillan, 1934.
- [11] J.-E. Littlewood G.-H. Hardy and G. Polya. *Inequalities*. Cambridge University Press, 1967.
- [12] C. Gentile and M. Warmuth. Proving relative loss bounds for on-line learning algorithms using Bregman divergences. In *Tutorials of the 13<sup>th</sup> International Conference on Computational Learning Theory*, 2000.
- [13] W. Leontief. *The Structure of the American Economy*. Harvard University Press, 1941.
- [14] E.-A. Mitscherlich. Das gesetz des minimums und das gesetz des abnehmenden bodenertrages. *Landw. Jahrb.*, 38:537–552, 1909.
- [15] F. Nielsen, J.-D. Boissonnat, and R. Nock. On Bregman Voronoi diagrams. In *Proc. of the 19<sup>th</sup> ACM-SIAM Symposium on Discrete Algorithms*, pages 746–755, 2007.
- [16] R. Nock and F. Nielsen. On Weighting Clustering. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28:1223–1235, 2006.
- [17] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- [18] W.-J. Spillman. *The law of diminishing returns*. World Book Co., 1924.
- [19] J.-H. von Thünen. *Der isolierte Staat in Beziehung auf Landwirthschaft und Nationalökonomie*. Fischer, 1930.

## 6 Appendix

### 6.1 Proof of Theorem 1

Theorem 4 in [3] states the result for a function that meets [1], [3] and  $F(\mathbf{x}, \mathbf{x}) = 0, \forall \mathbf{x} \in \mathbb{R}^d$  (the  $\Leftarrow$  in [2]).  $F$  also meets identity of indiscernibles as otherwise indeed, (2) would be violated for any distribution with distinct  $\mathbf{x}$  and  $\mathbf{y}$  as support and such that  $F(\mathbf{x}, \mathbf{y}) = 0$ . In this case, the solution of the left-hand side of (2) would be  $\mathbf{x}$  or  $\mathbf{y}$ , but not their average.

### 6.2 Proof of Theorem 3

Because of its independent interest for Bregman divergences, we state the proof in the most general form: we do not make the assumption that the generator is separable (4). Without loss of generality, we assume  $\Gamma = 1$  in (8). The concavity of  $\mu_\varphi$  means:

$$\mathbf{E}_j \nabla_\varphi^{-1}(\mathbf{E}_i \nabla_\varphi(\mathbf{x}_{ij})) \leq \nabla_\varphi^{-1}(\mathbf{E}_i \nabla_\varphi(\mathbf{E}_j \mathbf{x}_{ij})) . \quad (37)$$

Let  $\mathbf{x}_{ij} \stackrel{\text{def}}{=} \nabla_\varphi^{-1}(\mathbf{x}'_{ij})$  for  $\mathbf{x}'_{ij} \in \text{im}(\nabla_\varphi)$ . Applying  $\nabla_\varphi$  on both sides ( $\varphi$  is strictly convex, so  $\nabla_\varphi$  is bijective) and replacing yields:

$$\nabla_\varphi(\mathbf{E}_j \nabla_\varphi^{-1}(\mathbf{E}_i \mathbf{x}'_{ij})) \leq \mathbf{E}_i \nabla_\varphi(\mathbf{E}_j \nabla_\varphi^{-1}(\mathbf{x}'_{ij})) . \quad (38)$$

Eq. (38) states the convexity of the LDA  $\boldsymbol{\mu} \stackrel{\text{def}}{=} \nabla_{\varphi}(\mathbf{E}_{\mathcal{D}}\nabla_{\varphi}^{-1}(\mathbf{X}))$ , but Legendre duality implies  $\nabla_{\varphi} = \nabla_{\varphi^*}^{-1}$ , and we get  $\boldsymbol{\mu} = \boldsymbol{\mu}_{\varphi^*}$ , the dual of LDA  $\mu_{\varphi}$ . The proof starting from the convexity of  $\mu_{\varphi}$  follows the same path.

### 6.3 Proof of Theorem 4

We prove the first part of the Theorem relating  $\mu_{\varphi}$  and  $\mathbf{s}$ . Without loss of generality and to save notations, we make the proof for separable generators (4) and take a route slightly different from a direct use of Jensen's inequality. Furthermore, we make the proof for the concave case, assuming (8) holds for  $\mu_{\varphi}$ . Division by  $\Gamma$  does not change the regime, and so  $\mu_{\varphi}/\Gamma = \nabla_{\phi}^{-1}(\sum_{i=1}^m (\gamma_i/\Gamma)\nabla_{\phi}(x_i))$  is concave. We compute the tangential hyperplane to  $\mu_{\varphi}/\Gamma$  when  $x_i = \tilde{x} \in \text{int}(\text{dom}(\phi)), \forall i = 1, 2, \dots, m$ . We know that since  $\mu_{\varphi}/\Gamma$  is concave, it shall be located below this hyperplane. We have  $\nabla_{\mu_{\varphi}/\Gamma} = [\dots (\gamma_i/\Gamma)\phi^{[2]}(x_i)/\phi^{[2]}(\mu_{\varphi}) \dots]^{\top}$ , and so the tangential hyperplane to  $\tilde{\mathbf{x}} = [\dots \tilde{x} \dots]^{\top}$  on any  $\mathbf{x}$  is (with  $\kappa_i \stackrel{\text{def}}{=} \gamma_i/\Gamma$ ):

$$\begin{aligned} z(\mathbf{x}) &= \mu_{\varphi}(\tilde{\mathbf{x}}) + \langle \mathbf{x} - \tilde{\mathbf{x}}, \nabla_{\mu_{\varphi}/\Gamma}(\tilde{\mathbf{x}}) \rangle \\ &= \tilde{x} + \langle \boldsymbol{\kappa}, \mathbf{x} \rangle - \tilde{x} \\ &= \langle \boldsymbol{\kappa}, \mathbf{x} \rangle = \sum_{i=1}^m (\gamma_i/\Gamma)x_i . \end{aligned} \quad (39)$$

We obtain  $\mu_{\varphi}/\Gamma \leq \sum_{i=1}^m (\gamma_i/\Gamma)x_i$ . This yields to the statement that  $\mu_{\varphi} \leq \mathbf{s}$ , as claimed. The proof for the convex case is similar. The second part of the Theorem is an immediate consequence of the first part, and so this ends the proof of Theorem 4.

### 6.4 Proof of Lemma 1

We have  $\phi_{\text{CES}}^{[1]}(x) = a(2\sigma - 1)x^{(\sigma-1)/\sigma}/\sigma$  and  $\left(\phi_{\text{CES}}^{[1]}\right)^{-1}(x) = (\sigma x/(a(2\sigma - 1)))^{\sigma/(\sigma-1)}$ . There remains to remark that  $B = \sum_{i=1}^m \beta_i = B^{\frac{\sigma}{\sigma-1}}$ , use (8) and get:

$$\mu_{\phi_{\text{CES}}} = B^{\frac{\sigma}{\sigma-1}} \left( \frac{1}{B^{\frac{\sigma}{\sigma-1}}} \sum_{i=1}^m B^{\frac{1}{\sigma-1}} \beta_i x_i^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \quad (40)$$

which, after simplification, gives (10), as claimed.

### 6.5 Proof of Theorem 5

Without loss of generality, we first show that, if  $x_{\star}$  is a LDA that meets (21), it is a CES. We thus assume the form (8) for  $x_{\star}$ . Using (8), we obtain that  $x_{\star}$  satisfies (we replace in this proof notation  $\phi_x$  by the simpler  $\phi$  for the sake of readability):

$$\phi^{[1]}\left(\frac{x_{\star}}{\Gamma}\right) = \frac{1}{\Gamma} \sum_{i=1}^m \gamma_i \phi^{[1]}(x_i), \quad (41)$$

with  $\gamma_i \geq 0, \forall i = 1, 2, \dots, m$ , and  $\Gamma \stackrel{\text{def}}{=} \sum_{i=1}^m \gamma_i$ . If we differentiate (12) with respect to any  $x_i$ , using (8), we get:

$$\frac{z_i}{z_{\star}} = \frac{\gamma_i \phi^{[2]}(x_i)}{\phi^{[2]}(x_{\star}/\Gamma)}. \quad (42)$$

We multiply both sides by  $x_i/(\Gamma x_\star)$ , sum for all  $i$ , simplify via (12), rearrange, and get:

$$\frac{x_\star}{\Gamma} \phi^{[2]} \left( \frac{x_\star}{\Gamma} \right) = \frac{1}{\Gamma} \sum_{i=1}^m \gamma_i x_i \phi^{[2]}(x_i) . \quad (43)$$

Now, we match (41) with (43), and get that  $\phi$  *must* satisfy:

$$\exists \kappa \in \mathbb{R}_* \text{ s.t. } \phi^{[1]}(x) = \kappa x \phi^{[2]}(x), \forall x \in \text{dom} \phi . \quad (44)$$

The solution is found to be  $\phi^{[1]}(x) \propto x^\kappa$ , i.e.:

$$\phi(x) = \frac{d}{\kappa + 1} x^{\kappa+1} , \quad (45)$$

with  $d \in \mathbb{R}_*$  any constant that keeps (45) convex. Matching (45) with (11) implies  $\sigma = 1/(1 - \kappa)$ , and we get the proof that  $x_\star$  is a CES.

Lemma 2 then implies that  $z_\star$  is also a CES. The proof that  $\phi_z^{[2]} = d(\phi^{[2]})^{-1}$  for some  $d \in \mathbb{R}_*$  follows from the expressions of  $\phi_c$  and  $\phi_p$  in Lemmata 1 and 2.

## 6.6 Proof of Theorem 6

It is well-known that the property is true for any CES, so we investigate the reverse implication, and given any LDA  $x_\star$ , let  $\phi$  (which replaces  $\phi_x$  for the sake of readability) denote its generator. Using (8), we obtain:

$$E_{x_\star}^{x_i} = \frac{\gamma_i x_i \phi^{[2]}(x_i)}{x_\star \phi^{[2]}(x_\star/\Gamma)} . \quad (46)$$

Summing this for  $i = 1, 2, \dots, m$  and rearranging, we get that the sum of elasticities of  $x_\star$  is one implies:

$$\frac{x_\star}{\Gamma} \phi^{[2]} \left( \frac{x_\star}{\Gamma} \right) = \frac{1}{\Gamma} \sum_{i=1}^m \gamma_i x_i \phi^{[2]}(x_i) . \quad (47)$$

This is exactly (43). There remains to conclude that  $x_\star$  is a CES as in the proof of Theorem 5.

## 6.7 Proof of Theorem 7

Implication  $\Leftarrow$  is folklore, so we investigate the reverse implication. For any LDA  $x_\star$  whose generator is denoted  $\phi$ ,  $E_{x_\star}^{x_i \rightarrow x_j} = 1$  implies, with  $\tilde{x}_{ij} \stackrel{\text{def}}{=} x_j/x_i$ :

$$\frac{dS_{x_\star}^{x_i \rightarrow x_j}}{d\tilde{x}_{ij}} = \frac{S_{x_\star}^{x_i \rightarrow x_j}}{\tilde{x}_{ij}} , \quad (48)$$

implying  $S_{x_\star}^{x_i \rightarrow x_j} = \kappa \tilde{x}_{ij}$ , for some  $\kappa > 0$  which does not depend on  $x_i$  or  $x_j$ . We obtain that  $x_\star$  satisfies the following PDE:

$$x_i \left( \frac{\partial x_\star}{\partial x_i} \right) - \kappa x_j \left( \frac{\partial x_\star}{\partial x_j} \right) = 0 . \quad (49)$$

Because  $x_\star$  is a LDA, we have  $\partial x_\star / \partial x_i = \gamma_i \phi^{[2]}(x_i) / \phi^{[2]}(x_\star)$ , and so (49) becomes:

$$\gamma_i x_i \phi^{[2]}(x_i) = \kappa \gamma_j x_j \phi^{[2]}(x_j) . \quad (50)$$



Since  $\kappa, \gamma_i, \gamma_j > 0$  and (50) holds for any  $x_i, x_j \in \text{int}(\text{dom}\phi^{[1]})$ , we obtain that  $x\phi^{[2]}(x)$  is constant. This yields  $\phi(x) = b(x \log x - x)$  for some constant  $b \in \mathbb{R}_{+,*}$ , the generator of Cobb-Douglas LDA (18), as claimed.

*Remark:* (50) also proves that if  $m > 2$  and we require unit substitution elasticity between more than two goods, then necessarily  $\kappa = 1$  and  $\gamma_i = \gamma_j, \forall i, j = 1, 2, \dots, m$ . By means of words, fixing unit substitution elasticity for more than two goods implies equal weights for the goods. Clearly, this is a property of Cobb-Douglas aggregator rather than a restriction for LDAs.

## 6.8 Proof of Theorem 8

Implication  $\Leftarrow$  is folklore, so we investigate the reverse implication. For any LDA  $x_\star$  whose generator is denoted  $\phi$ , (25) implies:

$$\Gamma \nabla_\phi^{-1} \left( \frac{1}{\Gamma} \sum_{i=1}^m \gamma_i \nabla_\phi(\lambda x_i) \right) = \lambda^a \Gamma \nabla_\phi^{-1} \left( \frac{1}{\Gamma} \sum_{i=1}^m \gamma_i \nabla_\phi(x_i) \right). \quad (51)$$

Take some  $x_i, i = 1, 2, \dots, m$ , and differentiate both sides in  $x_i$ . We get after simplification:

$$\frac{\lambda \phi^{[2]}(\lambda x_i)}{\phi^{[2]}(\lambda^a x_\star)} = \frac{\lambda^a \phi^{[2]}(x_i)}{\phi^{[2]}(x_\star)}. \quad (52)$$

- Case 1:  $a \neq 1$ . Suppose that  $x_i = x_j = z/\lambda, \forall i, j = 1, 2, \dots, m$ , which implies  $x_\star = z/\lambda$  as well. Eq. (52) simplifies to:

$$\phi^{[2]}(\lambda^{a-1} z) = \frac{1}{\lambda^{a-1}} \phi^{[2]}(z), \quad (53)$$

i.e.  $\phi^{[2]}$  is homogeneous of degree  $-1$ . Euler's homogeneous function Theorem implies that  $\phi^{[2]}$  satisfies the following PDE:

$$x\phi^{[3]}(x) + \phi^{[2]}(x) = 0, \quad (54)$$

whose solution is  $\phi^{[2]}(x) \propto 1/x$  (with a positive factor), i.e.  $\phi(x) = b(x \log x - x)$  for some constant  $b \in \mathbb{R}_{+,*}$ , the generator of Cobb-Douglas LDA (18), as claimed.

- Case 2:  $a = 1$ . In this case, (52) implies:

$$\phi^{[2]}(\lambda x) = g(\lambda) \phi^{[2]}(x), \quad (55)$$

for any function  $g(\lambda) \in \mathbb{R}_*$ . Suppose without loss of generality that  $g$  is  $C_1$ , so that we can take the route of the proof of Euler's homogeneous function Theorem. We differentiate (55) in  $\lambda$ , and take the resulting equation for  $\lambda = 1$ . We obtain the following PDE:

$$x\phi^{[3]}(x) - g^{[1]}(1)\phi^{[2]}(x) = 0, \quad (56)$$

i.e.  $\phi^{[2]}(x) \propto x^\kappa$ , where  $\kappa \in \mathbb{R}_*$  is some constant. We obtain that  $\phi$  is either of the form of (11), or (18), the generators of CES and Cobb-Douglas LDAs, as claimed.

## 6.9 Proof of Lemma 4

After differentiation on some  $x_i$ , should it be a LDA, any MST aggregator  $x_\star$  with generator  $\phi$  would satisfy:

$$\gamma_i \times \frac{-\theta \exp(\theta \gamma_i x_i)}{(1 - \exp(\theta \gamma_i x_i))} \times x_\star = \gamma'_i \times \phi^{[2]}(x_i) \times \frac{1}{\phi^{[2]}(x_\star)}, \quad (57)$$

with  $\gamma'_i$  the LDA weight for  $x_i$ . This would imply  $\phi^{[2]}(x) = 1/x$ , from which the simplification of (57) yields that regardless of the value of  $x_i$ , the corresponding weights  $\gamma_i$  and  $\gamma'_i$  must satisfy  $-\theta \gamma_i x_i \exp(\theta \gamma_i x_i) = \gamma'_i (1 - \exp(\theta \gamma_i x_i))$ , impossible.

## 6.10 Proof of Theorem 9

Let  $\lambda$  be the Lagrange multiplier for (32), so that the Lagrangian is  $L \stackrel{\text{def}}{=} u_\star + \lambda(r - p_\star c_\star - m)$ , and we obtain the following stationarity conditions for the optimum:

$$\frac{\partial L}{\partial c_\star} = \frac{\gamma \phi_u^{[2]}(c_\star)}{\phi_u^{[2]}(u_\star)} - \lambda p_\star = 0, \quad (58)$$

$$\frac{\partial L}{\partial m} = \frac{(1 - \gamma) \phi_u^{[2]}(m/p_\star)}{p_\star \phi_u^{[2]}(u_\star)} - \lambda = 0. \quad (59)$$

Solving (59) for  $\lambda$  and simplifying (58) yields:

$$\phi_u^{[2]}(c_\star) = \frac{1 - \gamma}{\gamma} \phi_u^{[2]} \left( \frac{m}{p_\star} \right). \quad (60)$$

There remains to use the identity  $p_\star c_\star + m = r$  to get (34) and (35), as claimed.