

Information geometries for microeconomic theories

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

Brice Magdalou*

brice.magdalou@martinique.univ-ag.fr

Nicolas Sanz

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

Eric Briys*

eric.b@cyberlibris.com

Fred Célimène*

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

January 20, 2009

Abstract

More than thirty years ago, Charnes, Cooper and Schinnar (1976) established an enlightening contact between economic production functions (EPFs) — a cornerstone of neoclassical economics — and information theory, showing how a generalization of the Cobb-Douglas production function encodes homogeneous functions.

As expected by Charnes *et al.*, the contact turns out to be much broader: we show how information geometry as pioneered by Amari

and others underpins static and dynamic descriptions of microeconomic cornerstones.

We show that the most popular EPFs are fundamentally grounded in a very weak axiomatization of economic transition costs between inputs. The strength of this characterization is surprising, as it geometrically bonds altogether a wealth of collateral economic notions — advocating for applications in various economic fields —: among all, it characterizes (i) Marshallian and Hicksian demands and their geometric duality, (ii) Slutsky-type properties for the transformation paths, (iii) Roy-type properties for their elementary variations.

1 Introduction

Microeconomic theory builds from the behavior of individual agents — consumers and producers — to compute aggregate economic outcomes [16]. A cornerstone of the neoclassical school of economics consists in performing such aggregations using *Economic Production Functions* (EPFs). Obliterating the technical underpinnings [23], an EPF μ_x simply aggregates a set of inputs x_1, x_2, \dots, x_m :

$$\mu_x \doteq f(x_1, x_2, \dots, x_m) , \quad (1)$$

where μ_x is the output. We use the term “production” because of historical reasons [17, 23], but EPFs can be used to aggregate any economically relevant inputs, such as consumptions, prices, productions, labors, capitals, incomes, etc. [5, 12, 13]. As such, EPFs are not only used to model consumers’ and producers’ behaviors: they are virtually used in any branch of economic analysis [22], and even other fields as well [10, 11].

A significant part of microeconomic theory, grounded in the use of these EPFs, has been widely criticized for its *ad hoc* status, despite longstanding celebrated successes [24]. Accordingly, EPFs look like solutions in search of a problem. Our intention is not to settle the debate from a restrictive economics standpoint. It is rather to build upon an enlightening rationale [7], and follow further their information theoretic Ariadne’s thread to the core of microeconomic theories.

We start with a weak axiomatization of *economic transition costs* between different combinations of inputs, inspired by more recent works in information geometry [1, 3]. It has dramatic consequences: it grounds the most popular EPFs as optimal and exhaustive for transition costs from general standpoints; it provides an exhaustive description of *economic transformation paths* and the algorithmics of transition. The dynamics of transition

surprisingly reproduce properties well-known for EPFs (*e.g.* Slutsky's and Roy's identities), and patch the expansion path to new paths that we call *Hicksian* and *Marshallian*.

Section 2 presents preliminary definition and properties; section 3 gives our main results; a last section discusses and concludes. In order not to laden the paper's body, proofs are postponed to an appendix at the end of the paper.

2 Preliminary definitions and properties

2.1 Bregman divergences and LDAs

Bold notations such as \mathbf{x} denote vector-based notations, and blackboard faces such as \mathbb{X} sets of (tuples of) real numbers of \mathbb{R} or natural integers of \mathbb{N} . The information-theoretic part of this paper relies on two principal tools: Bregman divergences [6], and Lowest Distortion Aggregators.

Definition 1 *Let $\varphi : \mathbb{X} \rightarrow \mathbb{R}$ be strictly convex, differentiable over the interior of \mathbb{X} , with $\mathbb{X} \subseteq \mathbb{R}^d$ convex:*

- the Bregman Divergence D_φ with generator φ is [6, 3]:

$$D_\varphi(\mathbf{x}_i|\mathbf{x}_j) \doteq \varphi(\mathbf{x}_i) - \varphi(\mathbf{x}_j) - (\mathbf{x}_i - \mathbf{x}_j)^\top \nabla_\varphi(\mathbf{x}_j) , \quad (2)$$

where $\nabla_\varphi(\mathbf{x}_j) \doteq [\partial\varphi(\mathbf{x}_j)/\partial x_{jk}]^\top$ is the gradient operator;

- let $\mathbb{S} \doteq \{(\mathbf{x}_i, \gamma_i)\}_{i=1}^m$, with $\gamma_i \in \mathbb{R}_{+*}, i = 1, 2, \dots, m$. The Lowest Distortion Aggregator (LDA) μ_φ with generator φ for set \mathbb{S} is:

$$\mu_\varphi \doteq \nabla_\varphi^{-1} \left(\frac{1}{\Gamma} \sum_{i=1}^m \gamma_i \nabla_\varphi(\mathbf{x}_i) \right) , \Gamma \doteq \sum_{i=1}^m \gamma_i . \quad (3)$$

The shape of μ_φ is that of an Economic Production Function (EPF), in which each \mathbf{x}_i is a possibly multidimensional input. This is a setting far more general than mainstream economics where each x_i would be a scalar input (1), \mathbb{X} an interval of \mathbb{R} ($d = 1$), and so ∇_φ the conventional derivative. To distinguish this conventional case when $d = 1$, we shall use the shorthands for derivatives:

$$\varphi^{[k]}(x) \doteq d^k \varphi(x) / dx^k, \forall k \in \mathbb{N}_* , \quad (4)$$

$$U(x) \doteq \varphi^{[1]}(x) . \quad (5)$$

The use of (5) instead of the general (4) is intended to help the reader grasp potential LDA applications with utility functions: in economics of risk, μ_φ would be the certainty equivalent for function U [16]; in normative economics, μ_φ would be the equally distributed equivalent income for function U [4], and so on. Notice that multidimensionality is economically interesting, because in this setting each \mathbf{x}_i could represent a set of capital, a set of labour, etc. . However, for the sake of clarity and to remain stick to mainstream, most of the remaining of this paper, starting from the next subsection, is devoted to the 1-dimensional input case. Table 1 displays popular Bregman divergences tailored to this setting.

EPFs are grounded in baseline mathematical properties, such as concavity or convexity that play important roles in shaping preference or aversion for diversity [5, 12]. Following are baseline properties for LDAs that could be of use in an EPF setting. A part of the following Lemma is proofsketched in the Appendix.

Lemma 1 *The following properties hold true for any LDA μ_φ :*

- (*min - max bounds*) $\mathbf{x}_{\min} \leq \mu_\varphi \leq \mathbf{x}_{\max}$, where $x_{\min,j} = \min_i x_{ij}$ and $x_{\max,j} = \max_i x_{ij}, \forall j = 1, 2, \dots, m$;
- (*stability under composition*) the composition of LDAs with the same generator φ is a LDA with generator φ ;
- (*invariance modulo linear transforms*) let $\varphi_{\mathbf{b},c}(\mathbf{x}_i) \doteq \varphi(\mathbf{x}_i) + \mathbf{b}^\top \mathbf{x}_i + c$, with $\mathbf{b} \in \mathbb{R}^d, c \in \mathbb{R}$. Then $D_{\varphi_{\mathbf{b},c}}(\mathbf{x}_i || \mathbf{x}_j) = D_\varphi(\mathbf{x}_i || \mathbf{x}_j)$ and $\mu_{\varphi_{\mathbf{b},c}} = \mu_\varphi$;
- (*concavity - convexity duality*) μ_φ is concave if and only if its dual μ_{φ^*} is convex, where φ^* is the Legendre conjugate of φ :

$$\varphi^*(\mathbf{x}_i) \doteq \sup_{\mathbf{y} \in \mathbb{X}} \{\mathbf{y}^\top \mathbf{x}_i - \varphi(\mathbf{y})\} . \quad (6)$$

- (*relationship with arithmetic LDA*) if μ_φ is concave (resp. convex), then it is upperbounded (resp. lowerbounded) by the arithmetic LDA:

$$\mu \doteq \frac{1}{\Gamma} \sum_{i=1}^m \gamma_i \mathbf{x}_i . \quad (7)$$

2.2 LDAs and EPFs

We let p be the output price, and $p_i, i = 1, 2, \dots, m$ the price of input i , with $P \doteq \sum_{i=1}^m p_i$.

2.2.1 Marshallian and Hicksian demands on LDAs

The following Lemma (proof straightforward) gives an important result in connection with EPFs.

Lemma 2 *The expansion path of concave LDA μ_φ , i.e. the solution of $\max_{\{x_i\}_{i=1}^m} \{p\mu_\varphi - \sum_{i=1}^m p_i x_i\}$, is given by the sets of m -tuples $\{x_i\}_{i=1}^m$ such that:*

$$\varphi^{[2]}(x_i) = \frac{(p_i/\gamma_i)}{(p_j/\gamma_j)} \varphi^{[2]}(x_j), \forall i, j = 1, 2, \dots, m. \quad (8)$$

The equivalent problem can be formulated for convex LDAs, after flipping min/max.

This Lemma admits interesting economic consequences, some beyond the scope of this paper. Here is an example.

Lemma 3 *On the expansion path of any LDA μ_φ , the marginal rate of substitution of x_i for x_j (18) satisfies $S_{\mu_\varphi}^{x_i \rightarrow x_j} = (p_i/\gamma_i)/(p_j/\gamma_j)$, for any $i, j = 1, 2, \dots, m$.*

There are two important economic problems related to the optimization of EPFs: maximizing income under a budget constraint, and minimizing expenditures under an output constraint. Let us cast them for general LDAs.

Definition 2 *The Marshallian demand for concave LDA μ_φ is the problem:*

$$\begin{aligned} \{x_i\}_{i=1}^m &= \arg \max_{\{y_i\}_{i=1}^m} pU^{-1} \left(\frac{1}{\Gamma} \sum_{i=1}^m \gamma_i U(y_i) \right) \\ &\text{s.t. } \sum_{i=1}^m p_i y_i \leq w, \end{aligned} \quad (9)$$

The Hicksian demand for concave LDA μ_φ is the problem:

$$\begin{aligned} \{x_i\}_{i=1}^m &= \arg \min_{\{y_i\}_{i=1}^m} \sum_{i=1}^m p_i y_i \\ &\text{s.t. } pU^{-1} \left(\frac{1}{\Gamma} \sum_{i=1}^m \gamma_i U(y_i) \right) \geq p\mu'. \end{aligned} \quad (10)$$

Above, $w > 0$ is an income and $\mu' \geq 0$ an output level. Equivalent problems may be formulated for convex LDAs, after flipping min/max and the inequalities.

The proof of the following Lemma easily follows from Lemma 2. It is an important economic sanity check on LDAs.

Lemma 4 *For any concave (convex) LDA, the Marshallian and Hicksian demands belong to its expansion path.*

2.2.2 Price-Transition Cost Balanced setting

Lemma 2 says that the expansion path of some LDA is not necessarily linear in general. Yet, popular EPFs have linear expansion paths, such as Cobb-Douglas. Most remarkably, an important setting produces a linear subspace in the expansion path of *any* LDA. If one makes the assumption that p_i/γ_i is a constant for any $i = 1, 2, \dots, m$ — a situation to which we refer as *Price-Transition Cost Balanced* (PTCB for short, rationale in the following Section) —, then the \mathbb{R}^m linear subspace \mathbb{L} defined by $x_1 = x_2 = \dots = x_m$ inside the domain of μ_φ belongs to the expansion path of μ_φ . This is stated below.

Lemma 5 *The Marshallian and Hicksian demands in the PTCB setting are:*

$$x_j = \mu, \forall j = 1, 2, \dots, m \quad (\text{Marshallian}), \quad (11)$$

$$x_j = \mu_\varphi, \forall j = 1, 2, \dots, m \quad (\text{Hicksian}), \quad (12)$$

where μ is given in (7) and μ_φ is given in (3).

3 Main results

3.1 Mainstream EPFs are LDAs

A natural question on LDAs is whether they can accurately represent a significant part of mainstream EPFs. We answer affirmatively this question on the basis of the six categories of EPFs presented on the left column of Table 2 (Theorem also summarized in the Table).

Theorem 1 *The following holds true:*

- (I) *the following EPFs (Table 2) are LDAs (3):*
- *Constant Elasticity of Substitution (CES), for generator:*

$$\varphi_{\text{CES}}(x) \doteq ax^{2-\frac{1}{\sigma}} + bx + c; \quad (13)$$

- *Cobb-Douglas with constant returns to scale (CD, $\sum_i \beta_i = 1$), for generator:*

$$\varphi_{\text{CD}}(x) \doteq a'x \log x + bx + c ; \quad (14)$$

- *Generalized Exponential Mean (GEM), for generator:*

$$\varphi_{\text{GEM}}(x) \doteq a' \exp(\theta x + d) + bx + c . \quad (15)$$

(II) *Leontief is a limit case of LDA, for $\sigma \rightarrow 0^+$ in (13);*

(III) *Translog and MSTs are not LDAs.*

Above, $a \in \mathbb{R}_*$ is such that (13) is convex, $a' \in \mathbb{R}_{+*}$, $b, c, d \in \mathbb{R}$.

(proofsketch: see the Appendix). Thus, the most popular EPFs appearing in the theories of the consumer and producer, but also in normative economics, are in fact LDA[16, 4]. CES is the most favourably positioned: with the exception of Arimoto divergences, CES spans the LDAs corresponding to all Bregman divergences in Table 1.

3.2 Mainstream EPFs are economically exhaustive for LDAs

In this section, we switch to the main analytical economic assumptions that can be made about EPFs, and check which LDAs satisfy them. The (non-empty) subset of LDAs obtained is called “exhaustive” for the assumption. To distinguish between different sets of inputs, a general LDA/EPF for inputs x_1, x_2, \dots, x_m shall be denoted μ_x (φ is implicit) (1). We summarize these assumptions. The first defines *dually coupled* EPFs μ_x and μ_z , that satisfy:

$$\sum_{i=1}^m x_i z_i = \mu_x \mu_z . \quad (16)$$

(16) states that EPFs behave in the same way as their components, but at the highest (aggregation) level. Important examples include aggregating prices and consumptions, and aggregating wages and labor demands [5, 12]. The other assumptions rely on elasticities, substitution elasticities, marginal rates of substitutions, homogeneity and translatability:

$$E_{\mu_x}^{x_i} \doteq \left(\frac{d\mu_x}{\mu_x} \right) / \left(\frac{dx_i}{x_i} \right) \quad (17)$$

is the elasticity of μ_x with respect to x_i ,

$$S_{\mu_x}^{x_i \rightarrow x_j} \doteq \left(\frac{\partial \mu_x}{\partial x_i} \right) / \left(\frac{\partial \mu_x}{\partial x_j} \right) \quad (18)$$

is the marginal rate of substitution of x_i for x_j in μ_x , and

$$E_{\mu_x}^{x_i \rightarrow x_j} \doteq \left(\frac{d(x_j/x_i)}{x_j/x_i} \right) / \left(\frac{dS_{\mu_x}^{x_i \rightarrow x_j}}{S_{\mu_x}^{x_i \rightarrow x_j}} \right) \quad (19)$$

is the substitution elasticity of x_i for x_j in μ_x . Finally, function $f(x_1, x_2, \dots, x_m)$ is homogeneous of degree $a \in \mathbb{R}_*$ if and only if $(\forall \lambda)$:

$$f(\lambda x_1, \lambda x_2, \dots, \lambda x_m) = \lambda^a f(x_1, x_2, \dots, x_m) , \quad (20)$$

and translatable if and only if $(\forall \lambda)$:

$$f(\lambda + x_1, \lambda + x_2, \dots, \lambda + x_m) = \lambda + f(x_1, x_2, \dots, x_m) . \quad (21)$$

Theorem 2 *Let μ_x be any LDA. The following holds true:*

- (A) (16) holds if and only if μ_x and μ_z are CES;
- (B) $\sum_{i=1}^m E_{\mu_x}^{x_i} = 1$ if and only if μ_x is a CES;
- (C) $\exists 1 \leq i, j \leq m$ such that $E_{\mu_x}^{x_i \rightarrow x_j} = 1$ if and only if μ_x is a Cobb-Douglas with constant returns to scale;
- (D) $\exists 1 \leq i, j \leq m$ such that $E_{\mu_x}^{x_i \rightarrow x_j} = a \in \mathbb{R}_{+*}$ if and only if μ_x is a CES;
- (E) μ_x is homogeneous of degree 1 if and only if it is a CES;
- (F) μ_x is translatable if and only if it is a GEM.

(proof: see the Appendix). Since Cobb-Douglas with constant returns to scale is a particular case of CES, it meets simultaneously the assumptions in **A-E**. Theorem 2 says that if one casts LDAs into an analytical setting compatible with mainstream economic assumptions, then the huge set of LDAs reduces to mainstream economic EPFs. This, we think, is a clear-cut position of LDAs in favor of their economic “viability”. The rightmost columns of Table 2 summarize the results of Theorem 2 for each couple (assumption, EPF), using symbols Y(es), N(o) or L = in the limit.

3.3 An axiomatization of global transitions costs

A natural question is to quantify the global transition cost $\mathcal{F}(\mathbf{x}, \mathbf{y})$ when input x_i shifts to y_i (for $i = 1, 2, \dots, m$). Assume \mathcal{F} is separable, *i.e.* sums input-based contributions:

$$\mathcal{F}(\mathbf{x}, \mathbf{y}) \doteq \sum_{i=1}^m \gamma_i F(x_i, y_i) , \quad (22)$$

where F quantifies the distortion of input i during the transition, while $\gamma_i > 0$ is its relative Unit Transition Cost (hereafter UTC), and we use the shorthand vector notation $\mathbf{x} \doteq (x_1, x_2, \dots, x_m)$. Our goal is to exhibit light conditions on both parameters to fully specify the economic transition, and by the way ground LDAs as meeting optimality conditions that parallel those of Marshallian and Hicksian demands (9, 10). We start by γ_i , and make the assumption that γ_i is proportional to p_i , *i.e.* the UTC for input i is proportional to its unit price. This is the Price-Transition Cost Balanced setting (PTCB). We make three assumptions on F . The two first are structural: **(i)** it is non negative, *i.e.* it is lower-bounded; **(ii)** the local transition cost for input i is zero if and only if both inputs are the same, *i.e.* if and only if $x_i = y_i$. The third assumption is economic: **(iii)** specifies the average input value which minimizes \mathcal{F} (22); more precisely, this average is just the average inputs leveraged by their respective prices. This last assumption connects input prices to UTCs, and justifies the PTCB setting.

Quite remarkably, these assumptions on UTCs and F are necessary *and* sufficient to completely shape the setting for economic costs *and* transitions.

Theorem 3 *In the PTCB setting, assume that F is twice continuously differentiable and meets the following assumptions:*

- (i)** *non-negativity:* $F(x_i, y_i) \geq 0$;
- (ii)** *identity of indiscernibles:* $F(x_i, y_i) = 0 \Leftrightarrow x_i = y_i$;
- (iii)** *the inputs average minimizes the global transition cost:*

$$\arg \min_y \frac{1}{\Gamma} \sum_{i=1}^m \gamma_i F(x_i, y) = \frac{1}{P} \sum_{i=1}^m p_i x_i . \quad (23)$$

Then $F(x, y) = D_\varphi(x||y)$ for some strictly convex and differentiable φ .

(The proof is a slight variation to that of Theorem 4 in [3]). Any Bregman divergence satisfies (1.), (2.) and (3.), and so the characterization of F in Theorem 3 is almost exhaustive given the mild regularity conditions imposed. Let us review the main consequences of Theorem 3. Hereafter, we sometimes use the shorthand $w \doteq \sum_{i=1}^m p_i x_i$.

3.4 Transition costs top EPFs and their dual geometries

Most remarkably, the assumptions of Theorem 3 yields that the arithmetic average is not the only smallest global transition cost: since D_φ is not necessarily symmetric, we may also compute the solution to (23) in which x_i and y are switched. The solution comes naturally as Legendre duality enters the analysis, as we have [19]:

$$D_\varphi(x_i||y) = D_{\varphi^*}(U(y)||U(x_i)) , \quad (24)$$

and so the optimum sought immediately follows:

$$\arg \min_y \sum_{i=1}^m \gamma_i D_\varphi(y||x_i) = \mu_\varphi . \quad (25)$$

This is just the LDA as formulated in a more general setting in (3), and justifies the name ‘‘LDA’’.

From an economic perspective of the PTCB setting, any transition cost grounds an optimal EPF (the LDA) which defines lowest cost transitions (its linear expansion path, Lemma 5). The Hicksian demand emerges as a geometric dual (25) of the Marshallian demand (23), a consequence of the property that transition costs already define dual affine geometries for the inputs, in the economic input space (x_i) and in its image by U (5). This connection is well studied in differential information geometry [1], and completes the popular economic duality between both demands [16].

3.5 Transition costs underlie economic transformation paths

We now move onto a less static description of the transition, and show that global transition costs (22) are integrals computed over a particular *economic transformation path* between \mathbf{x} and \mathbf{y} . Due to its importance, the Theorem to come is given in the most general setting: d is arbitrary (Definition 1), and the divergence is not assumed to be separable [6].

Definition 3 Let φ be a function meeting the conditions of Definition 1. The matching loss parametrized by φ is the path integral for vector field $\mathbf{v}_\varphi(\mathbf{z})$:

$$G(\mathbf{x}_i, \mathbf{y}_i) \doteq \int_{\mathbb{P}} \mathbf{v}_\varphi^\top(\mathbf{z}) d\mathbf{z} , \quad (26)$$

where the vector field and the path are respectively:

$$\mathbf{v}_\varphi(\mathbf{z}) \doteq \nabla_\varphi(\mathbf{z}) - \nabla_\varphi(\mathbf{x}_i) , \quad (27)$$

$$\mathbb{P} \doteq \{ \mathbf{z}(\lambda) \doteq (1 - \lambda)\mathbf{x}_i + \lambda\mathbf{y}_i, \lambda \in [0, 1] \} . \quad (28)$$

The link between the path integral and Bregman divergences is stated in the following Theorem.

Theorem 4 $G(\mathbf{x}_i, \mathbf{y}_i) = D_\varphi(\mathbf{y}_i || \mathbf{x}_i)$.

(proof: see the Appendix). Theorem 4 tells us that the economic transition is a linear transformation from \mathbf{x}_i to \mathbf{y}_i . Figure 1 provides us with an especially interesting example of transition path directly mapped on the EPF in the PTCB setting (left, in grey; the EPF is a Cobb-Douglas). This path goes to some input state $\mathbf{x} \doteq (x_1, x_2)$, starting from its Marshallian demand on the expansion path (this is point $\boldsymbol{\mu} \doteq (\mu, \mu)$). Let us denote this grey path on the EPF which links $\boldsymbol{\mu}$ to \mathbf{x} the *Marshallian path* of \mathbf{x} . As we move along the Marshallian path in the economic input space, we move along a curve on the dual mean, which goes to $U(\mathbf{x})$, starting from its Hicksian demand $U(\boldsymbol{\mu}_\varphi)$. We call this dual path, which follows an isoquant of the dual mean, an *Hicksian path* (right picture in Figure 1). We can also define equivalently Hicksian paths in the economic input space, and Marshallian paths on the dual mean. To make an analogy with physical string deformations, the vector field (27) (Figure 1, right) is just the force required to keep distorted a string with one endpoint fixed at $U(\boldsymbol{\mu}_\varphi)$, and the other endpoint somewhere along the Hicksian path. The economic transition cost is thus analogous to a *work*.

3.6 Slutsky-type transformations

We show that *any* transformation can be equivalently decomposed in *two* transformations. This decomposition bears surprising similarities with those involved in a fundamental microeconomic equation, Slutsky's identity [16]. Our starting point is the following identity, elsewhere known as Bregman triangle equality [20] (for any $\mathbf{x}, \mathbf{y}, \mathbf{z}$ in $\text{dom}(\varphi)^m$):

$$\mathcal{F}(\mathbf{y}, \mathbf{x}) = \mathcal{F}(\mathbf{y}, \mathbf{z}) + \mathcal{F}(\mathbf{z}, \mathbf{x}) + \Delta(\mathbf{x}, \mathbf{y}, \mathbf{z}) , \quad (29)$$

where $\Delta(\mathbf{x}, \mathbf{y}, \mathbf{z}) \doteq \sum_{i=1}^m \gamma_i(x_i - z_i)(u(z_i) - u(y_i))$. For any \mathbf{x} and \mathbf{y} in $\text{dom}(\varphi)^m$, there always exist a \mathbf{z} in $\text{dom}(\varphi)^m$ for which $\Delta(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 0$ [1]. This \mathbf{z} , which we call the Bregman-Slutsky Inputs (BSI) of \mathbf{x} and \mathbf{y} , yields a decomposition of $\mathcal{F}(\mathbf{x}, \mathbf{y})$ in two transitions costs. Figure 2 presents an example of this decomposition.

The similarity with the decomposition of Slutsky's identity is striking, yet the framework of (29) is much different: Slutsky's identity decomposes the variation of the Marshallian demand when input prices change. Hence, the output is completely specified by the change in prices, while (29) assumes absolutely nothing about the reasons for the output's change. According to Slutsky's identity, the change in the output can be decomposed in a substitution effect between inputs, and an income effect. (29) tells us similar facts when one endpoint of the transformation is along the expansion path in the PTCB setting (proof immediate from (24) and (29)).

Theorem 5 $\forall c, c' \in \text{dom}(\varphi)$,

$$\sum_{i=1}^m \gamma_i D_\varphi(x_i | c) = \sum_{i=1}^m \gamma_i D_\varphi(x_i | \mu) + \Gamma D_\varphi(\mu | c) , \quad (30)$$

$$\sum_{i=1}^m \gamma_i D_\varphi(c' | x_i) = \Gamma D_\varphi(c' | \mu_\varphi) + \sum_{i=1}^m \gamma_i D_\varphi(\mu_\varphi | x_i) . \quad (31)$$

Hence, in the PTCB setting, when one endpoint of the transformation lies on the expansion path, the corresponding BSI is the Marshallian or Hicksian demand, also on the expansion path. Figure 3 presents the two types of transitions, from and to the expansion path: the transformation from \mathbf{x} to \mathbf{c}' on the expansion path (yellow) is the composition of two paths:

- (a) from \mathbf{x} to its Hicksian demand, on the Hicksian path, and
- (b) from this Hicksian demand to \mathbf{c}' , on the expansion path.

(a) is no more than substitution effect on the transition cost, and (b) the income effect on the transition cost. (31) tells us that leaving the expansion path trades the substitution effect for a *budget* effect in the second stage.

3.7 Roy-type elementary variations

Roy's identity is also fundamental in microeconomics; it says that provided μ_φ meets mild assumptions, we have $S_{\mu_\varphi}^{p_i \rightarrow w} = x_i$ [16] (18). It is not hard to prove that *any* elementary input distortion, to or from μ_φ , meets Roy's identity when μ_φ does (proofsketch in Appendix).

Theorem 6 We have ($\forall c \in \text{dom}(\varphi)$):

$$S_{D_\varphi(c|\mu_\varphi)}^{p_i \rightarrow w} = S_{D_\varphi(\mu_\varphi|c)}^{p_i \rightarrow w} = S_{\mu_\varphi}^{p_i \rightarrow w} . \quad (32)$$

4 Discussion and conclusion

The anticipations of Charnes *et al.* [7] are finally not surprising: a significant part of the neoclassical school of economics is about formalizing and aggregating information, and so bonds with information theory had to be expected. What is striking is that those bonds for EPFs come from a weak characterization of — moreover — a completely different standpoint on aggregation (dynamic, with geometric flavors). Furthermore, the transition standpoint rejoins quantities and properties popular in the “static” (EPF) cases — Marshallian and Hicksian demands, Slutsky’s and Roy’s identities, to name a few.

There are various general follow-ups to expect from such a work to continue upon [7], two of which appear to be particularly interesting from the economic standpoint. First, an alternative to the transition cost (22) is to compute the maximum cost over inputs. What is economically interesting is that the population minimizer (minmax) is *still* the general LDA (3) but on a combinatorial basis of at most $d + 1$ inputs [21]: the leveraging coefficients are not the γ_i s anymore, at most $d + 1$ of them are $\neq 0$, they do not have a closed form, but they admit an efficient approximation algorithm [21]. Second, alleviating the constraint $d = 1$, and even the separability of the global transition cost (22) leads to a rich economic setting of interactions between inputs that deserves further studies.

References

- [1] S.-I. Amari and H. Nagaoka. *Methods of Information Geometry*. Oxford University Press, 2000.
- [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] C. Blackorby, W. Bossert, and D. Donaldson. Income inequality measurement: the normative approach. In J. Silber, editor, *Handbook of income inequality measurement*, pages 133–157. Kluwer, 1999.
- [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] A. Charnes, W.-W. Cooper, and A.-P. Schinnar. A theorem on homogeneous functions and extended Cobb-Douglas forms. *Proceedings of the National Academy of Sciences USA*, 73:3747–3748, 1976.
- [8] L.-R. Christensen, D.-W. Jorgenson, and L.-J. Lau. Conjugate duality and the transcendental logarithmic production function. *Econometrica*, 39:255–256, 1971.
- [9] C. Cobb and P.-H. Douglas. A theory of production. *American Economic Review*, 18:139–165, 1928.
- [10] G. Conway. Mathematical models in applied ecology. *Nature*, 269:291–297, 1977.
- [11] H. de Kroon, A. Plaisier, J. van Groenendael, and H. Caswell. Elasticity: the relative contribution of demographic parameters to population growth rate. *Ecology*, 67:1427–1431, 1986.
- [12] A.-K. Dixit and J.-E. Stiglitz. Monopolistic competition and optimum product diversity. *American Economic Review*, 67:297–308, 1977.
- [13] R. Griffin, J. Montgomery, and M. Rister. Selecting functional form in production function analysis. *Western Journal of Agricultural Economics*, 12:216–227, 1987.
- [14] W. Leontief. *The Structure of the American Economy*. Harvard University Press, 1941.
- [15] F. Liese and I. Vajda. *Convex Statistical Distances*. Teubner Verlag, 1987.
- [16] A. Mas-Colell, M. Winston, and J. Green. *Microeconomic Theory*. Oxford University Press, 1995.

- [17] S.-K. Mishra. A brief history of production functions, 2007. <http://mpira.ub.uni-muenchen.de/5254/>.
- [18] E.-A. Mitscherlich. Das gesetz des minimums und das gesetz des abnehmenden bodenertrages. *Landw. Jahrb.*, 38:537–552, 1909.
- [19] F. Nielsen, J.-D. Boissonnat, and R. Nock. On Bregman Voronoi diagrams. In *Proc. of the 19th ACM-SIAM Symposium on Discrete Algorithms*, pages 746–755. SIAM - ACM Press, 2007.
- [20] R. Nock, P. Luosto, and J. Kivinen. Mixed Bregman clustering with approximation guarantees. In *Proc. of the 23rd European Conference on Machine Learning*, pages 154–169. Springer-Verlag, 2008.
- [21] R. Nock and F. Nielsen. Fitting the Smallest Enclosing Bregman Ball. In *Proc. of the 20th European Conference on Machine Learning*, pages 649–656. Springer-Verlag, 2005.
- [22] A. Shaikh. Non-linear dynamics and pseudo-production functions. *Eastern Economic Journal*, 31:447–466, 2005.
- [23] R. W. Shephard. *Theory of Costs and Production Functions*. Princeton University Press, 1970.
- [24] R.-M. Solow. Technical change and the aggregate production function. *Review of Economics and Statistics*, 39:312–320, 1957.
- [25] W.-J. Spillman. *The law of diminishing returns*. World Book Co., 1924.
- [26] J.-H. von Thünen. *Der isolierte Staat in Beziehung auf Landwirthschaft und Nationalökonomie*. Fischer, 1930.

5 Appendix: proofs

5.1 Proofs sketch of Lemma 1, fourth point

Without loss of generality, we assume $\Gamma = 1$ in (3). Using the mathematical expectation notation \mathbf{E} in lieu of the average to save space, the concavity of μ_φ means $\mathbf{E}_j \nabla_\varphi^{-1}(\mathbf{E}_i \nabla_\varphi(\mathbf{x}_{ij})) \leq \nabla_\varphi^{-1}(\mathbf{E}_i \nabla_\varphi(\mathbf{E}_j \mathbf{x}_{ij}))$. Let $\mathbf{x}_{ij} \doteq \nabla_\varphi^{-1}(\mathbf{x}'_{ij})$ for $\mathbf{x}'_{ij} \in \text{im}(\nabla_\varphi)$. Applying ∇_φ on both sides (φ is strictly convex, so ∇_φ is bijective) and replacing yields:

$$\nabla_\varphi(\mathbf{E}_j \nabla_\varphi^{-1}(\mathbf{E}_i \mathbf{x}'_{ij})) \leq \mathbf{E}_i \nabla_\varphi(\mathbf{E}_j \nabla_\varphi^{-1}(\mathbf{x}'_{ij})) . \quad (33)$$

Eq. (33) states the convexity of the LDA $\tilde{\boldsymbol{\mu}} \doteq \nabla_{\varphi}(\mathbf{E}\nabla_{\varphi}^{-1}(\mathbf{X}))$, but Legendre duality implies $\nabla_{\varphi} = \nabla_{\varphi^*}^{-1}$, and we get $\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu}_{\varphi^*}$, the dual of LDA $\boldsymbol{\mu}_{\varphi}$. The proof starting from the convexity of $\boldsymbol{\mu}_{\varphi}$ follows the same path.

5.2 Profsketch of Theorem 1

We only treat the case of MST (point **(III)**). We differentiate the MST in x_i . If it is a LDA with generator φ , μ_x must satisfy:

$$\gamma_i \times \frac{-\theta \exp(\theta \gamma_i x_i)}{(1 - \exp(\theta \gamma_i x_i))} \times \mu_x = \gamma'_i \times \varphi^{[2]}(x_i) \times \frac{1}{\varphi^{[2]}(\mu_x)}, \quad (34)$$

with γ'_i the LDA weight for x_i . Looking at μ_x , this would imply $\varphi^{[2]}(x) = 1/x$, from which the simplification of (34) yields that regardless of the value of x_i , the corresponding weights γ_i and γ'_i must satisfy $-\theta \gamma_i x_i \exp(\theta \gamma_i x_i) = \gamma'_i (1 - \exp(\theta \gamma_i x_i))$, impossible.

5.3 Proof of Theorem 2

For all six equivalences, implication \Leftarrow is folklore. We prove the reverse implications for points **(A)**, **(E)** and **(F)**. The remaining proofs exploit the same tools.

(point **(A)**). Since μ_x is a LDA with generator φ , it satisfies:

$$\varphi^{[1]}(\mu_x) = \frac{1}{\Gamma} \sum_{i=1}^m \gamma_i \varphi^{[1]}(x_i). \quad (35)$$

We differentiate (16) in x_i , use (3), and get:

$$\frac{z_i}{\mu_z} = \frac{\gamma_i \varphi^{[2]}(x_i)}{\varphi^{[2]}(\mu_x)}. \quad (36)$$

We multiply both sides by x_i , sum for all i , simplify via (16), rearrange, and get:

$$\mu_x \varphi^{[2]}(\mu_x) = \frac{1}{\Gamma} \sum_{i=1}^m \gamma_i x_i \varphi^{[2]}(x_i). \quad (37)$$

Now, we match (35) with (37), and get that φ must satisfy:

$$\exists \kappa \in \mathbb{R}_* \text{ s.t. } \varphi^{[1]}(x) = \kappa x \varphi^{[2]}(x), \forall x \in \text{dom} \varphi. \quad (38)$$

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

$$\varphi(x) = \frac{d}{\kappa + 1} x^{\kappa+1}, \quad (39)$$

with $d \in \mathbb{R}_*$ any constant that keeps (39) convex. Matching (39) with (13) implies $\sigma = 1/(1 - \kappa)$, and we get the proof that μ_x is a CES. The other EPF, μ_z , can be found by inspecting (16) after remarking that partial derivatives on the left and right-hand side must also coincide. After a standard derivation using the general CES form for μ_x (Table 2), we obtain that μ_z is:

$$\mu_z = \left(\sum_{i=1}^m \beta_i^\sigma z_i^{1-\sigma} \right)^{\frac{1}{1-\sigma}}, \quad (40)$$

which is also a CES, with generator $\varphi(x) = bx^{2-\sigma}$, with $b \in \mathbb{R}_*$ any constant for which φ is convex.

(point **(E)**). Consider some LDA μ_x whose generator is denoted φ . Without losing too much generality, we assume that φ is twice continuously differentiable, as in Theorem 3. (20) implies:

$$\begin{aligned} & \left(\varphi^{[1]} \right)^{-1} \left(\frac{1}{\Gamma} \sum_{i=1}^m \gamma_i \varphi^{[1]}(\lambda x_i) \right) \\ &= \lambda^a \left(\varphi^{[1]} \right)^{-1} \left(\frac{1}{\Gamma} \sum_{i=1}^m \gamma_i \varphi^{[1]}(x_i) \right). \end{aligned} \quad (41)$$

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

$$\frac{\lambda \varphi^{[2]}(\lambda x_i)}{\varphi^{[2]}(\lambda^a \mu_x)} = \frac{\lambda^a \varphi^{[2]}(x_i)}{\varphi^{[2]}(\mu_x)}. \quad (42)$$

(42) implies:

$$\varphi^{[2]}(\lambda x) = g(\lambda) \varphi^{[2]}(x), \quad (43)$$

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 (43) in λ , and take the resulting equation for $\lambda = 1$. We obtain the following PDE:

$$x \varphi^{[3]}(x) - g^{[1]}(1) \varphi^{[2]}(x) = 0, \quad (44)$$

i.e. $\varphi^{[2]}(x) \propto x^\kappa$, where $\kappa \in \mathbb{R}_*$ is some constant. We obtain that φ is either of the form of (13), or (14), the generators of CES or Cobb-Douglas with constant returns to scale (a particular case of CES), as claimed.

(point **(F)**). We take the same route as for the proof of point **(E)**, but differentiating (21) instead of (20). (42) becomes after simplification:

$$\frac{\varphi^{[2]}(\lambda + x_i)}{\varphi^{[2]}(x_i)} = \frac{\varphi^{[2]}(\lambda + \mu_x)}{\varphi^{[2]}(\mu_x)} , \quad (45)$$

which yields $\varphi^{[2]}(x + \lambda) = g(\lambda)\varphi^{[2]}(x)$ for some function g , and, taking $x = 0$, brings $g(\lambda) = \varphi^{[2]}(\lambda)/\varphi^{[2]}(0)$. We obtain $\varphi^{[2]}(x + \lambda) = \varphi^{[2]}(\lambda)\varphi^{[2]}(x)/\varphi^{[2]}(0)$, implying $\varphi^{[2]}(x) \neq 0$, and thus $\varphi^{[2]}(2x) = (\varphi^{[2]})^2(x)/\varphi^{[2]}(0)$. The change of variable $g(x) \doteq \ln(|\varphi^{[2]}(x)|)$ yields:

$$g(2x) = 2g(x) - g(0) , \quad (46)$$

and thus $g^{[1]}(2x) = g^{[1]}(x) = K$, a constant. Taking the route back to $\varphi^{[1]}$, we easily obtain:

$$\varphi^{[1]}(x) = a \exp(\theta x + b) + c , a \in \mathbb{R}_{+*}, \theta \in \mathbb{R}_*, b, c \in \mathbb{R} ,$$

from which we recover the LDA of the Generalized Exponential Mean (Table 2).

5.4 Proof of Theorem 4

One easily recovers Bregman divergences as $d\mathbf{z} = (\mathbf{y}_i - \mathbf{x}_i)d\lambda$, and the integral becomes:

$$\begin{aligned} G(\mathbf{x}_i, \mathbf{y}_i) &= \int_0^1 (\mathbf{y}_i - \mathbf{x}_i)^\top (\nabla_\varphi(\mathbf{z}(\lambda)) - \nabla_\varphi(\mathbf{x}_i)) d\lambda & (47) \\ &= -(\mathbf{y}_i - \mathbf{x}_i)^\top \nabla_\varphi(\mathbf{x}_i) \\ &\quad + \int_0^1 (\mathbf{y}_i - \mathbf{x}_i)^\top \nabla_\varphi((1 - \lambda)\mathbf{x}_i + \lambda\mathbf{y}_i) d\lambda \\ &= -(\mathbf{y}_i - \mathbf{x}_i)^\top \nabla_\varphi(\mathbf{x}_i) + \int_0^1 d\varphi((1 - \lambda)\mathbf{x}_i + \lambda\mathbf{y}_i) \\ &= -(\mathbf{y}_i - \mathbf{x}_i)^\top \nabla_\varphi(\mathbf{x}_i) + \varphi(\mathbf{y}_i) - \varphi(\mathbf{x}_i) \\ &= D_\varphi(\mathbf{y}_i || \mathbf{x}_i) . & (48) \end{aligned}$$

5.5 Proofs sketch of Theorem 6

The proof is immediate once we remark that the derivative of the distortion is proportional to the derivative of the LDA:

$$\frac{\partial D_\varphi(c|\mu_\varphi)}{\partial u} = -\varphi^{[2]}(\mu_\varphi)(c - \mu_\varphi) \frac{\partial \mu_\varphi}{\partial u}, \quad (49)$$

$$\frac{\partial D_\varphi(\mu_\varphi|c)}{\partial u} = (\varphi^{[1]}(\mu_\varphi) - \varphi^{[1]}(c)) \frac{\partial \mu_\varphi}{\partial u}. \quad (50)$$

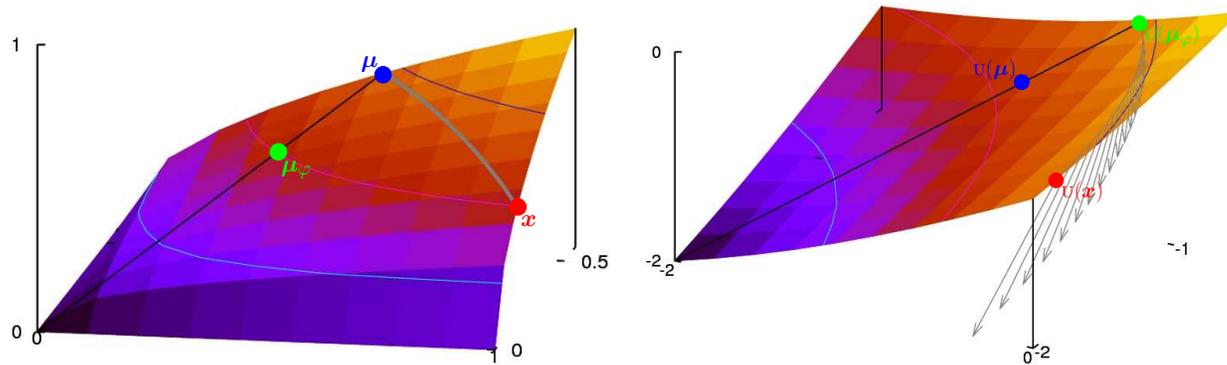


Figure 1: Left: a Cobb-Douglas ($m = 2$); right: its dual (GEM with $\theta = 1$, Table 2). Isoquants are plotted for both LDAs. On the left, inputs displayed on the LDA are $\mathbf{x} \doteq (x_1, x_2)$, $\boldsymbol{\mu}_\varphi \doteq (\mu_\varphi, \mu_\varphi)$, $\boldsymbol{\mu} \doteq (\mu, \mu)$. The grey paths are the Marshallian (left) and Hicksian (right) paths; the black lines on the LDAs are their expansion paths.

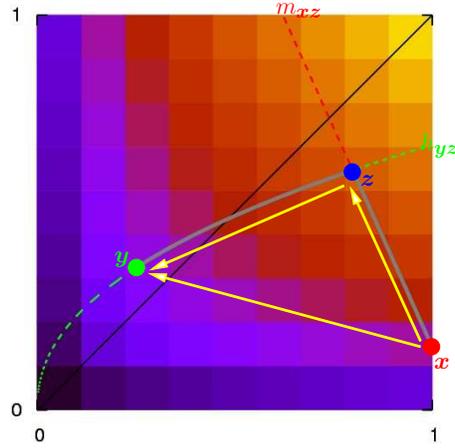


Figure 2: The economic transformation path from \mathbf{x} to \mathbf{y} can always be decomposed in two (29), involving a particular set of inputs \mathbf{z} , the BSI of \mathbf{x} and \mathbf{y} (yellow arrows; the EPF is a Cobb-Douglas and $m = 2$). Dashed curves $m_{\mathbf{xz}}$ and $h_{\mathbf{yz}}$ are orthogonal in that *any* transition from $m_{\mathbf{xz}}$ to $h_{\mathbf{yz}}$ admits \mathbf{z} as BSI (see text for details).

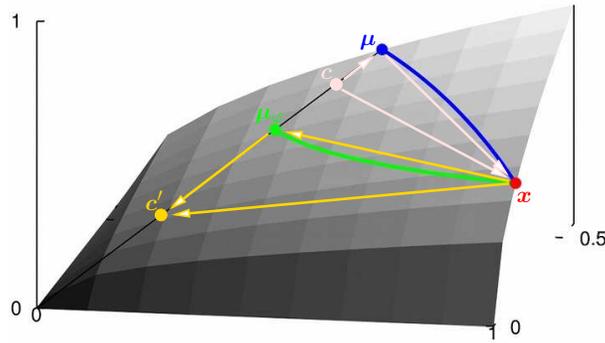


Figure 3: On the Cobb-Douglas LDA, $\mathbf{c} \doteq (c, c)$ and $\mathbf{c}' \doteq (c', c')$ are two points on the expansion path of the LDA. The transition costs, from \mathbf{c}' to some point $\mathbf{x} \doteq (x_1, x_2)$, or from \mathbf{x} to \mathbf{c} , can be exactly decomposed using only the expansion (black), Marshallian (blue) and Hicksian (green) paths (see text for details, and Figure 1 for the notations).

Table 1: Some popular Bregman divergences $D_\varphi(x_i||y_i)$.

$\text{dom}(\varphi)$	$\varphi(x)$	$D_\varphi(x_i y_i)$	Divergence name
\mathbb{R}	x^2	$(x_i - y_i)^2$	Squared Euclidean norm
\mathbb{R}_+	$x \log x - x$	$x_i \log \frac{x_i}{y_i} - x_i + y_i$	Kullback-Leibler divergence
\mathbb{R}_{+*}	$-\log x$	$\frac{x_i}{y_i} - \log \frac{x_i}{y_i} - 1$	Itakura-Saito divergence
$\mathbb{R}, \alpha \in (-1, 1)$	$\frac{4}{1-\alpha^2} \left(x - x^{\frac{1+\alpha}{2}} \right)$	$\frac{2}{1+\alpha} x_i^{\frac{1+\alpha}{2}} + \frac{2}{1-\alpha} y_i x_i^{\frac{\alpha-1}{2}} - \frac{4}{1-\alpha^2} y_i^{\frac{1+\alpha}{2}}$	Amari α -divergence
$\mathbb{R}_{+*}, \alpha \rightarrow -1$		=Kullback-Leibler divergence($x_i y_i$)	
$\mathbb{R}_+, \alpha \rightarrow 1$		=Kullback-Leibler divergence($y_i x_i$)	
$\mathbb{R}, \alpha \in (0, 1)$	$\frac{-x^\alpha + \alpha x - \alpha + 1}{\alpha(1-\alpha)}$	$\frac{1}{\alpha(1-\alpha)}(x_i^\alpha - y_i^\alpha - \alpha y_i x_i - \alpha x_i^2)$	Bregman-Csiszár divergence
$\mathbb{R}_{+*}, \alpha \rightarrow 0$		=Itakura-Saito divergence($x_i y_i$)	
$\mathbb{R}_+, \alpha \rightarrow 1$		=Kullback-Leibler divergence($x_i y_i$)	
$\mathbb{R}, \alpha \in (0, 1)$	$\frac{(x^{\frac{1}{\alpha}} + 1)^\alpha - 2^\alpha}{2(1-\alpha)}$	$\frac{1}{2(1-\alpha)} \left((x_i^{\frac{1}{\alpha}} + 1)^\alpha - (y_i^{\frac{1}{\alpha}} + 1)^{\alpha-1} (1 - x_i y_i^{\frac{1}{\alpha}-1} + 2 y_i^{\frac{1}{\alpha}}) \right)$	Arimoto divergence [15]
$\mathbb{R}, \alpha \rightarrow 0$		= F_1 divergence($x_i y_i$)	
$\mathbb{R}_{+*}, \alpha \rightarrow 1$		=Bayesian divergence($x_i y_i$)	

Table 2: Famous economic production functions $(\beta_i, \beta_{ij} > 0, \forall i, j)$, along with a summary of our results in Theorems 1 and 2 (see text for details). MSTs are given up to some eventual variable change.

EPF μ_x	Name	Optimality (Th. 1)	Exhaustivity (Th. 2)					
			(A)	(B)	(C)	(D)	(E)	(F)
$\left(\sum_{i=1}^m \beta_i x_i^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$	Constant Elasticity of Substitution (CES, $\sigma \in \mathbb{R}_* \setminus \{1\}$) [2]	Y	Y	Y	N	Y	Y	N
$\prod_{i=1}^m x_i^{\beta_i}$	Cobb-Douglas with constant returns to scale ($\sum_i \beta_i = 1$) [9, 16]	Y	Y	Y	Y	Y	Y	N
$(1/\theta) \log(\sum_{i=1}^m \beta_i \exp(\theta x_i))$	Generalized Exponential Mean (GEM, $\theta \in \mathbb{R}_*$) [4]	Y	N	N	N	N	N	Y
$\min_i \{\beta_i x_i\}$	Leontief [14]	L	L	L	N	L	L	N
$\exp(\beta_0 + \sum_{i=1}^m \beta_i \log x_i + \sum_{i=1}^m \sum_{j=1}^m \beta_{ij} \log x_i \log x_j)$	Translog [8]	N	N	N	N	N	N	N
$\prod_{i=1}^m (1 - \exp(\theta \beta_i x_i))$	Mitscherlich-Spillman-von Thünen (MST, $\theta \in \{-1, +1\}$) [18, 25, 26]	N	N	N	N	N	N	N