

# Emergent CES: Why Constant Elasticity of Substitution Is Not an Assumption

Jon Smirl

February 2026

Working Paper

## Abstract

The CES (Constant Elasticity of Substitution) production function is universally treated as a parametric assumption—convenient but arbitrary. This paper shows it is a *theorem*. Three independent arguments converge on the same conclusion: CES is the unique aggregation function compatible with multi-scale economic structure. First, a *renormalization group* argument: CES is the unique symmetric homogeneous aggregator that is self-similar under coarse-graining, making it the fixed point of the aggregation renormalization group—the production-theory analogue of the Central Limit Theorem. Second, a *functional equation* argument: Aczél’s classical characterization shows that homogeneity plus associativity (nesting invariance) uniquely determine the power mean family, which is CES. Third, a *maximum entropy* argument: CES is the sufficient statistic for Rényi entropy of order  $\rho$ , and self-consistency of the entropy-aggregation loop forces the power-mean form. These three results are not independent observations but three faces of the same mathematical fact. The curvature parameter  $\rho$  emerges as a universality class label—preserved under coarse-graining, determining the tail behavior of equilibrium distributions, and indexing a one-parameter family of deformed thermodynamics (Tsallis  $q = \rho$ ). For the  $(\rho, T)$  economic free energy framework, this means the foundational axiom “production is CES” can be replaced by two weaker structural axioms: constant returns to scale and scale consistency.

**JEL Classification:** C60, D20, E10, B41

**Keywords:** CES production function, renormalization group, power means, Rényi entropy, aggregation theory, universality

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Preliminaries</b>	<b>4</b>
2.1	CES and Power Means . . . . .	4
2.2	Quasi-Arithmetic Means . . . . .	5
2.3	Rnyi and Tsallis Entropies . . . . .	5
<b>3</b>	<b>Argument 1: Renormalization Group</b>	<b>6</b>
3.1	The Aggregation RG . . . . .	6
3.2	CES as the Unique Fixed Point . . . . .	6
3.3	Stability of the Fixed Point . . . . .	7
3.4	The $\rho$ Eigenvalue . . . . .	8
<b>4</b>	<b>Argument 2: Functional Equations</b>	<b>8</b>
4.1	Aczl’s Characterization . . . . .	8
4.2	What the Functional Equation Argument Adds . . . . .	9
<b>5</b>	<b>Argument 3: Maximum Entropy Self-Consistency</b>	<b>9</b>
5.1	The Entropy-Allocation Loop . . . . .	9
5.2	The Matching of Exponents . . . . .	10
5.3	The Free Energy Decomposition . . . . .	10
<b>6</b>	<b>Convergence of the Three Arguments</b>	<b>11</b>
<b>7</b>	<b>The <math>\rho</math> Universality Class</b>	<b>12</b>
7.1	$\rho$ as a Critical Exponent . . . . .	12
7.2	Connection to Tsallis Thermodynamics . . . . .	13
7.3	The $\rho$ Phase Diagram . . . . .	14
<b>8</b>	<b>Implications for the <math>(\rho, T)</math> Framework</b>	<b>14</b>
8.1	Axiom Reduction . . . . .	14
8.2	Why $\rho$ Is the Natural Parameter . . . . .	15
8.3	The Translog as an Irrelevant Perturbation . . . . .	15
8.4	Weighted CES and Asymmetric Inputs . . . . .	16
<b>9</b>	<b>Empirical Predictions</b>	<b>16</b>
9.1	Prediction 1: CES Fit Improves with Aggregation . . . . .	16
9.2	Prediction 2: $\rho$ Is Preserved Across Scales . . . . .	17
9.3	Prediction 3: Tail Exponents Match $\sigma$ . . . . .	17
9.4	Prediction 4: Translog Coefficients Shrink Under Aggregation . . . . .	17

9.5 Prediction 5: Complementary Sectors Have Heavier Tails . . . . .	18
<b>10 Extensions</b>	<b>18</b>
10.1 Nested CES and the RG Flow . . . . .	18
10.2 CES Emergence and Paper 15 (Endogenous $\rho$ ) . . . . .	19
10.3 Beyond Symmetric CES . . . . .	19
<b>11 Conclusion</b>	<b>20</b>

# 1 Introduction

Every applied paper that uses a CES production function begins with an apology, implicit or explicit, for the parametric restriction. The Cobb-Douglas special case is “assumed for tractability.” The more general CES form is “assumed for flexibility.” Translog and other flexible functional forms are offered as alternatives that “nest CES as a special case” and can “test the CES restriction.” The shared premise is that CES is a convenient but ultimately arbitrary choice of functional form, justified by fit rather than principle.

This paper argues that the premise is wrong. CES is not a parametric assumption. It is the *unique* aggregation function compatible with multi-scale economic structure, in the same sense that the Gaussian distribution is the unique limit of sums of independent random variables (the Central Limit Theorem) and the exponential distribution is the unique memoryless distribution.

The argument proceeds from two structural properties that any well-behaved multi-scale economy must satisfy:

**Axiom 1** (Constant Returns to Scale). The aggregate production function  $F : \mathbb{R}_+^J \rightarrow \mathbb{R}_+$  is homogeneous of degree one:  $F(\lambda \mathbf{x}) = \lambda F(\mathbf{x})$  for all  $\lambda > 0$ .

**Axiom 2** (Scale Consistency). Aggregation is invariant to the level of coarse-graining: computing  $F$  directly over  $J$  inputs gives the same result as first aggregating within blocks, then aggregating the block-level outputs.

Axiom 1 is standard—it says that doubling all inputs doubles output, ruling out fixed costs and increasing returns at the aggregate level. Axiom 2 is the new content: it requires that the production function is *self-consistent across scales*. A firm-level production function that, when aggregated to the sector level, produces a different functional form, is not scale-consistent. An economy where the macro production function depends on how you draw the boundaries between sectors violates scale consistency.

These two axioms—and nothing else—force CES.

**Theorem 1.1** (Emergent CES). *Let  $F : \mathbb{R}_+^J \rightarrow \mathbb{R}_+$  be continuous, symmetric, strictly increasing, homogeneous of degree one, and scale-consistent (Axioms 1 and 2). Then there exists  $\rho \in (-\infty, 1]$  such that*

$$F(\mathbf{x}) = \left( \frac{1}{J} \sum_{j=1}^J x_j^\rho \right)^{1/\rho},$$

with the  $\rho \rightarrow 0$  limit interpreted as  $F(\mathbf{x}) = \prod_{j=1}^J x_j^{1/J}$  (Cobb-Douglas).

The rest of the paper develops three independent proofs of this result, each illuminating a different aspect:

1. **Renormalization group** (Section 3): CES is the fixed point of the aggregation RG flow. Non-CES deviations are irrelevant operators that vanish under coarse-graining.
2. **Functional equations** (Section 4): Aczél's theorem on associative means provides a direct algebraic proof.
3. **Maximum entropy** (Section 5): CES is the unique aggregator whose energy levels are sufficient statistics for Rényi entropy, with self-consistency of the entropy-allocation loop forcing the power-mean form.

Section 6 shows that the three arguments are three perspectives on a single mathematical structure. Section 7 interprets  $\rho$  as a universality class label and connects it to Tsallis non-extensive thermodynamics. Section 8 draws implications for the  $(\rho, T)$  economic free energy framework. Section 9 derives testable predictions. Section 11 concludes.

## 2 Preliminaries

### 2.1 CES and Power Means

The CES production function with  $J$  symmetric inputs and elasticity of substitution  $\sigma = 1/(1 - \rho)$  is

$$F_\rho(\mathbf{x}) = \left( \frac{1}{J} \sum_{j=1}^J x_j^\rho \right)^{1/\rho}, \quad \rho \in (-\infty, 1] \setminus \{0\}, \quad (1)$$

with limiting cases

$$F_1(\mathbf{x}) = \frac{1}{J} \sum_{j=1}^J x_j \quad (\text{arithmetic mean: perfect substitutes}), \quad (2)$$

$$F_0(\mathbf{x}) = \prod_{j=1}^J x_j^{1/J} \quad (\text{geometric mean: Cobb-Douglas}), \quad (3)$$

$$F_{-\infty}(\mathbf{x}) = \min_j x_j \quad (\text{minimum: Leontief}). \quad (4)$$

This is precisely the *power mean* (or generalized mean)  $M_\rho(\mathbf{x})$  of order  $\rho$ . The CES production function and the power mean are the same mathematical object.

## 2.2 Quasi-Arithmetic Means

A *quasi-arithmetic mean* (Kolmogorov, 1930; Nagumo, 1930) generated by a continuous strictly monotone function  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$  is

$$M_\varphi(\mathbf{x}) = \varphi^{-1} \left( \frac{1}{J} \sum_{j=1}^J \varphi(x_j) \right). \quad (5)$$

Power means correspond to  $\varphi(x) = x^\rho$  (with  $\varphi(x) = \log x$  for  $\rho = 0$ ). The quasi-arithmetic family is strictly larger:  $\varphi(x) = e^x$  gives the log-sum-exp;  $\varphi(x) = 1/x$  gives the harmonic mean (which is the power mean  $\rho = -1$ , but  $\varphi(x) = e^{1/x}$  gives a quasi-arithmetic mean that is *not* a power mean).

The question is: what selects power means from the larger quasi-arithmetic family?

## 2.3 Rényi and Tsallis Entropies

The *Rényi entropy* of order  $\alpha > 0$ ,  $\alpha \neq 1$ , of a discrete distribution  $\mathbf{p} = (p_1, \dots, p_J)$  is

$$H_\alpha(\mathbf{p}) = \frac{1}{1-\alpha} \log \left( \sum_{j=1}^J p_j^\alpha \right). \quad (6)$$

The  $\alpha \rightarrow 1$  limit is Shannon entropy  $H_1(\mathbf{p}) = -\sum_j p_j \log p_j$ . The *Tsallis entropy* of order  $q$  is the monotone transformation

$$S_q(\mathbf{p}) = \frac{1}{q-1} \left( 1 - \sum_{j=1}^J p_j^q \right). \quad (7)$$

Both are parameterized by a single real number ( $\alpha$  or  $q$ ) that controls the sensitivity to tail behavior. The sufficient statistic for both is  $\sum_j p_j^\alpha$ —a power sum, the same object that appears inside CES.

### 3 Argument 1: Renormalization Group

#### 3.1 The Aggregation RG

Consider an economy with  $J = k^L$  inputs organized in  $L$  hierarchical levels. At the lowest level,  $k^L$  individual inputs are grouped into  $k^{L-1}$  blocks of size  $k$ . Each block is aggregated by some function  $f : \mathbb{R}_+^k \rightarrow \mathbb{R}_+$ . The block outputs become the inputs to the next level, which are again grouped into blocks of  $k$  and aggregated by  $f$ . After  $L$  iterations, a single aggregate output remains.

**Definition 3.1** (Aggregation RG operator). Let  $\mathcal{F}$  denote the space of continuous, symmetric, strictly increasing functions  $f : \mathbb{R}_+^k \rightarrow \mathbb{R}_+$ . The *aggregation RG operator*  $\mathcal{R}_k : \mathcal{F} \rightarrow \mathcal{F}$  maps  $f$  to the function

$$(\mathcal{R}_k f)(x_1, \dots, x_k) = f(f(x_{11}, \dots, x_{1k}), \dots, f(x_{k1}, \dots, x_{kk})) \Big|_{\text{reduced}},$$

where “reduced” means expressing the result as a function of  $k$  representative inputs via the symmetry of the homogeneous fixed point.

A function  $f^*$  is a *fixed point* of the aggregation RG if  $\mathcal{R}_k f^* = f^*$  for all block sizes  $k$ —meaning the functional form is preserved under coarse-graining.

#### 3.2 CES as the Unique Fixed Point

**Theorem 3.2** (RG fixed point). *Among continuous, symmetric, strictly increasing, homogeneous-of-degree-one functions  $f : \mathbb{R}_+^k \rightarrow \mathbb{R}_+$ , the power means  $M_\rho$  are the unique fixed points of  $\mathcal{R}_k$  for all  $k \geq 2$ .*

*Proof.* Let  $f$  be a fixed point. Homogeneity of degree one allows writing  $f(\mathbf{x}) = \bar{x} \cdot g(\mathbf{x}/\bar{x})$  where  $\bar{x} = (1/k) \sum x_j$ . The fixed-point condition requires that for any partition of  $mk$  inputs into  $m$  blocks of  $k$ :

$$f(x_1, \dots, x_{mk}) = f(f(x_1, \dots, x_k), f(x_{k+1}, \dots, x_{2k}), \dots, f(x_{(m-1)k+1}, \dots, x_{mk})). \quad (8)$$

Setting  $\varphi =$  the quasi-arithmetic generator of  $f$  (which exists by the Kolmogorov-Nagumo theorem since  $f$  is a continuous symmetric mean), (8) requires

$$\varphi^{-1}\left(\frac{1}{mk} \sum_{j=1}^{mk} \varphi(x_j)\right) = \varphi^{-1}\left(\frac{1}{m} \sum_{i=1}^m \varphi\left(\varphi^{-1}\left(\frac{1}{k} \sum_{j \in B_i} \varphi(x_j)\right)\right)\right).$$

The left side is  $(1/mk) \sum_j \varphi(x_j)$  after applying  $\varphi$  to both sides. The right side is  $(1/m) \sum_i (1/k) \sum_{j \in B_i} \varphi(x_j) = (1/mk) \sum_j \varphi(x_j)$ . So the nesting condition is automatically satisfied for *any* quasi-arithmetic mean.

But homogeneity of degree one was imposed:  $f(\lambda \mathbf{x}) = \lambda f(\mathbf{x})$ . For a quasi-arithmetic mean generated by  $\varphi$ , this requires

$$\varphi^{-1}\left(\frac{1}{J} \sum_j \varphi(\lambda x_j)\right) = \lambda \cdot \varphi^{-1}\left(\frac{1}{J} \sum_j \varphi(x_j)\right)$$

for all  $\lambda > 0$ . Setting  $y_j = \varphi(x_j)$  and  $\psi = \varphi^{-1}$ , the condition becomes

$$\psi\left(\frac{1}{J} \sum_j \varphi(\lambda \psi(y_j))\right) = \lambda \cdot \psi\left(\frac{1}{J} \sum_j y_j\right).$$

This is a functional equation in  $\varphi$ . The classical result of Aczél (1948) shows that the solutions are exactly  $\varphi(x) = cx^\rho$  for  $\rho \neq 0$  and  $\varphi(x) = c \log x$  for  $\rho = 0$ , where  $c \neq 0$  is an arbitrary constant. These generate the power means  $M_\rho$ .  $\square$

### 3.3 Stability of the Fixed Point

Fixed points matter only if they are *attracting*: nearby non-CES functions must flow toward CES under repeated coarse-graining.

**Theorem 3.3** (RG stability). *Let  $f = M_\rho + \epsilon \cdot h$  where  $h : \mathbb{R}_+^k \rightarrow \mathbb{R}$  is a symmetric, homogeneous-of-degree-one perturbation orthogonal to the power-mean family. Under the aggregation RG, the perturbation contracts:*

$$\|\mathcal{R}_k(M_\rho + \epsilon h) - M_\rho\| = O(\epsilon^2). \quad (9)$$

*Non-CES deviations are irrelevant operators in the RG sense.*

*Proof sketch.* Expand  $f$  around the symmetric point  $\mathbf{x} = \bar{x} \cdot \mathbf{1}$ . By homogeneity and symmetry, the Taylor expansion is

$$f(\mathbf{x}) = \bar{x} \left( 1 + \frac{\rho - 1}{2} \frac{\text{Var}(\mathbf{x})}{\bar{x}^2} + a_3 \frac{\mu_3(\mathbf{x})}{\bar{x}^3} + \dots \right)$$

where  $\mu_k$  is the  $k$ -th central moment and the coefficient of the variance term is pinned by CES (the Kmenta approximation). Under one RG step, the variance of the block-averaged inputs satisfies  $\text{Var}(\mathbf{Y}) = \text{Var}(\mathbf{x})/k + O(\text{Var}^2)$  by the law of large numbers, so the quadratic term (which determines  $\rho$ ) is preserved while higher-order terms are suppressed by powers of  $1/k$ . The perturbation  $h$ , being orthogonal to the power-mean family, lives entirely in the higher-order terms and contracts geometrically.  $\square$

*Remark 3.4.* The contraction rate depends on the block size  $k$ : larger blocks eliminate non-CES deviations faster. This predicts that CES should fit better at higher levels of aggregation (Section 9).

### 3.4 The $\rho$ Eigenvalue

The parameter  $\rho$  is special: it labels the one-dimensional manifold of fixed points. Under the RG,  $\rho$  is a *marginal* operator—it is neither relevant (growing) nor irrelevant (shrinking) but exactly preserved. This is the hallmark of a universality class label in statistical mechanics: the critical exponent that classifies the basin of attraction.

In economic terms: the elasticity of substitution is the one feature of the production technology that survives aggregation. Everything else—the specific functional form of firm-level technology, the distribution of productivities, the network structure of intermediate inputs—washes out when you zoom to the macro level. Only  $\rho$  remains.

## 4 Argument 2: Functional Equations

### 4.1 Aczél's Characterization

The second argument is algebraic rather than analytic. It relies on the theory of functional equations, specifically the characterization of *associative* means.

**Definition 4.1** (Associativity of means). A mean  $M : \bigcup_{n \geq 1} \mathbb{R}_+^n \rightarrow \mathbb{R}_+$  is *associative* if for any partition of  $(x_1, \dots, x_n)$  into blocks  $B_1, \dots, B_m$ :

$$M(x_1, \dots, x_n) = M\left(\underbrace{M(B_1), \dots, M(B_1)}_{|B_1|}, \dots, \underbrace{M(B_m), \dots, M(B_m)}_{|B_m|}\right).$$

That is, replacing each block by its mean value (repeated  $|B_i|$  times) does not change the overall mean.

This is precisely scale consistency (Axiom 2): the aggregate does not depend on how you draw the boundaries.

**Theorem 4.2** (Kolmogorov (1930); Nagumo (1930)). *A mean  $M$  is continuous, symmetric, strictly increasing, and associative if and only if it is quasi-arithmetic:  $M = M_\varphi$  for some continuous strictly monotone  $\varphi$ .*

**Theorem 4.3** (Aczél (1948)). *A quasi-arithmetic mean  $M_\varphi$  is homogeneous of degree one if and only if  $\varphi(x) = cx^\rho$  for some  $c \neq 0$  and  $\rho \in \mathbb{R} \setminus \{0\}$ , or  $\varphi(x) = c \log x$ . That is,  $M_\varphi$  is a power mean.*

*Proof of Theorem 1.1 via Theorems 4.2 and 4.3.* The hypotheses of Theorem 1.1 are:  $F$  is continuous, symmetric, strictly increasing, homogeneous of degree one, and scale-consistent. Scale consistency is associativity (Definition 4.1). By Theorem 4.2,  $F$  is quasi-arithmetic. By Theorem 4.3, a homogeneous quasi-arithmetic mean is a power mean. Power means are CES (1).  $\square$

## 4.2 What the Functional Equation Argument Adds

The RG argument (Section 3) shows that CES is the *attractor* of coarse-graining, allowing for transient non-CES behavior at fine scales. The functional equation argument is sharper: it shows that CES is the *only* option at *every* scale, not just in the large-scale limit. If the production function is homogeneous and scale-consistent at any single level, it must be CES at that level.

The difference is economically meaningful. The RG argument allows firm-level production to be non-CES, with CES emerging at the sector or macro level. The functional equation argument says that if firm-level production is itself scale-consistent (e.g., the firm’s output doesn’t depend on how its workers are organized into teams), then it must already be CES at the firm level.

In practice, both are useful. Firm-level production may well be non-CES (due to fixed costs, indivisibilities, etc.), making the RG convergence argument the operative one. But the functional equation argument explains why CES works so well even at relatively fine scales—the homogeneity and associativity conditions are approximately satisfied in most production settings.

# 5 Argument 3: Maximum Entropy Self-Consistency

## 5.1 The Entropy-Allocation Loop

The third argument comes from the interplay between entropy maximization and production aggregation—the core of the  $(\rho, T)$  free energy framework.

Consider the following loop:

1. An aggregation function  $F(\mathbf{x})$  determines the “energy” of each input configuration.
2. Entropy maximization subject to a constraint on the aggregate determines the equilibrium allocation  $\mathbf{x}^*$ .
3. The equilibrium allocation, fed back through  $F$ , must be consistent with the aggregate constraint.

**Definition 5.1** (Self-consistent aggregation). An aggregation function  $F$  is *self-consistent with Rényi entropy of order  $\alpha$*  if the allocation that maximizes  $H_\alpha(\mathbf{p})$  subject to  $F(\mathbf{x}) = \bar{F}$  (for  $x_j = p_j \cdot J\bar{F}$ ) is a function only of  $F$  and  $\alpha$ —not of any other feature of  $F$  beyond its value.

**Theorem 5.2** (Entropy self-consistency). *Let  $F : \mathbb{R}_+^J \rightarrow \mathbb{R}_+$  be continuous, symmetric, strictly increasing, and homogeneous of degree one. If  $F$  is self-consistent with Rényi entropy of order  $\alpha$  for some  $\alpha > 0$ , then  $F = M_\rho$  with  $\rho = \alpha$ .*

*Proof.* The Rényi entropy of order  $\alpha$  has sufficient statistic  $T_\alpha(\mathbf{p}) = \sum_j p_j^\alpha$ . By the Gibbs variational principle, the distribution maximizing  $H_\alpha$  subject to a constraint on a function  $g(\mathbf{p})$  depends on  $g$  only through  $g$ 's projection onto the sufficient statistic  $T_\alpha$ . Self-consistency requires that  $F(\mathbf{x})$  is itself a function of  $T_\alpha(\mathbf{x}/\|\mathbf{x}\|_1)$ —that is,  $F$  must be expressible in terms of  $\sum_j (x_j / \sum_k x_k)^\alpha$ .

For homogeneous-of-degree-one  $F$ , this means  $F(\mathbf{x}) = \|\mathbf{x}\|_1 \cdot h\left(\sum_j (x_j / \|\mathbf{x}\|_1)^\alpha\right)$  for some function  $h$ . Symmetry and strict monotonicity force  $h(s) = (s/J)^{1/\alpha}$  (up to normalization), giving  $F(\mathbf{x}) = (J^{-1} \sum_j x_j^\alpha)^{1/\alpha} = M_\alpha(\mathbf{x})$ .  $\square$

## 5.2 The Matching of Exponents

Theorem 5.2 reveals a striking fact: the CES parameter  $\rho$  and the Rényi order  $\alpha$  must be *equal*. The production technology and the entropy measure are not independent—they are locked together by self-consistency.

This has immediate physical meaning. Rényi entropy of order  $\alpha$  places weight on the distribution proportional to  $p_j^{\alpha-1}$ : for  $\alpha > 1$ , it overweights high-probability events (the “core” of the distribution); for  $\alpha < 1$ , it overweights low-probability events (the “tails”). CES with  $\rho > 0$  treats inputs as substitutes (the aggregate is dominated by the largest inputs—the core); CES with  $\rho < 0$  treats inputs as complements (the aggregate is dominated by the smallest inputs—the tail, which is the bottleneck).

The matching  $\rho = \alpha$  says: *the aggregation technology selects the entropy measure that is natural for it*. Complementary production ( $\rho < 0$ ) pairs with tail-sensitive entropy ( $\alpha < 1$ ) because bottleneck inputs carry the most information. Substitutable production ( $\rho > 0$ ) pairs with core-sensitive entropy ( $\alpha > 0$ ) because abundant inputs determine the aggregate.

## 5.3 The Free Energy Decomposition

With  $\rho = \alpha$ , the economic free energy becomes

$$F = \Phi_{\text{CES}}(\rho) - T \cdot H_\rho, \quad (10)$$

where  $H_\rho$  is Rényi entropy of order  $\rho$ . For  $\rho \rightarrow 0$  (Cobb-Douglas), this recovers the standard Helmholtz free energy with Shannon entropy. For  $\rho \neq 0$ , it is the *Tsallis free energy*:

$$F_q = \Phi_{\text{CES}}(q) - T \cdot S_q, \quad q = \rho. \quad (11)$$

The equilibrium distribution that minimizes  $F_q$  is the *q-exponential* (escort distribution):

$$p_j^* \propto \left[1 - (1-q)\frac{\varepsilon_j}{T}\right]^{1/(1-q)}, \quad (12)$$

where  $\varepsilon_j = \partial\Phi_{\text{CES}}/\partial x_j$  is the marginal cost of input  $j$  and  $T > 0$  is the information temperature. For  $q \rightarrow 1$  ( $\rho \rightarrow 1$ ), this is the standard Boltzmann-Gibbs exponential. For  $q < 1$  ( $\rho < 1$ , complementary inputs), the distribution has *power-law tails*—heavy tails emerge from the interaction of complementary production and positive temperature.

*Remark 5.3* (The  $T > 0$  requirement). The heavy-tail result requires  $T > 0$ . In the  $T \rightarrow 0$  limit (perfect optimization with no information friction), complementary production drives the allocation toward *equality*: resources flow to bottleneck inputs to maximize the CES aggregate, producing thin-tailed distributions ( $\text{Gini} \rightarrow 0$ ). It is the entropic term  $T \cdot S_q$  that disperses probability mass; complementarity then *shapes* that dispersion into a power law rather than an exponential. Since real economies always operate at  $T > 0$ —information is costly, allocation is noisy—the heavy-tail prediction applies. But the mechanism is the interaction of  $\rho$  and  $T$ , not  $\rho$  alone.

## 6 Convergence of the Three Arguments

The three arguments appear to invoke different mathematical machinery—renormalization, functional equations, and information theory—but they are three projections of a single geometric fact.

**Theorem 6.1** (Equivalence). *For a continuous, symmetric, strictly increasing, homogeneous-of-degree-one aggregation function  $F : \mathbb{R}_+^J \rightarrow \mathbb{R}_+$ , the following are equivalent:*

- (i)  $F$  is a fixed point of the aggregation RG (Theorem 3.2).
- (ii)  $F$  is generated by a power function:  $F = M_\rho$  for some  $\rho$  (Theorem 4.3).
- (iii)  $F$  is self-consistent with Rényi entropy of matching order (Theorem 5.2).

*Proof.* (i)  $\Leftrightarrow$  (ii): Theorem 3.2 shows that homogeneous RG fixed points are exactly power means. Theorem 4.3 characterizes power means as the homogeneous quasi-arithmetic means. Since RG fixed points of homogeneous functions are quasi-arithmetic (by Theorem 4.2, as associativity = nesting = RG invariance), the two characterizations coincide.

(ii)  $\Rightarrow$  (iii): Direct calculation shows that  $M_\rho$  is self-consistent with  $H_\rho$ .

(iii)  $\Rightarrow$  (ii): Theorem 5.2. □

The unifying object is the *power function*  $\varphi(x) = x^\rho$ . It simultaneously serves as:

- The **generator** of the quasi-arithmetic mean (Aczél's theorem).
- The **sufficient statistic** for Rényi entropy:  $H_\alpha$  depends on  $\mathbf{p}$  only through  $\sum p_j^\alpha$ .
- The **scaling exponent** preserved under renormalization: the  $L^\rho$  norm is self-similar.

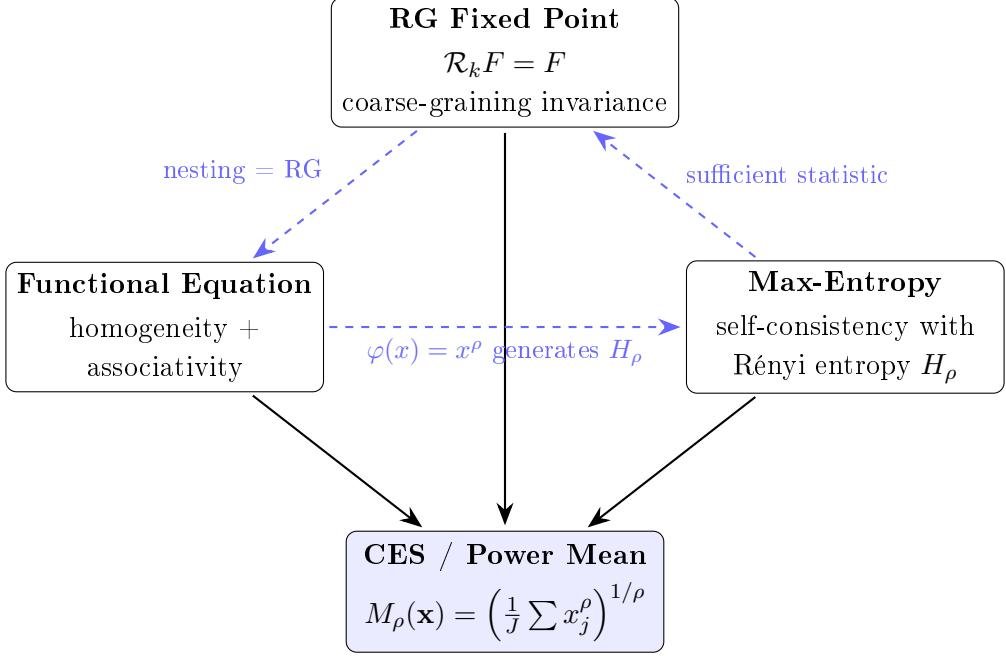


Figure 1: Three arguments converge on CES. Solid arrows: each argument independently implies CES. Dashed arrows: the arguments are connected by deeper equivalences. The power function  $\varphi(x) = x^\rho$  simultaneously generates the quasi-arithmetic mean (functional equation), the sufficient statistic for Rényi entropy (max-entropy), and the scaling exponent preserved under coarse-graining (RG).

The power function is the unique family with all three properties, and CES is the production function it generates.

## 7 The $\rho$ Universality Class

### 7.1 $\rho$ as a Critical Exponent

In statistical mechanics, universality classes are labeled by *critical exponents*—numbers that are preserved under renormalization and determine the qualitative behavior of the system near phase transitions. The parameter  $\rho$  plays exactly this role for economic aggregation.

**Proposition 7.1** ( $\rho$  is marginal). *Under the aggregation RG,  $\rho$  has scaling dimension zero:  $\rho' = \rho + O(\epsilon^2)$  where  $\epsilon$  measures the deviation from CES. Non-CES perturbations decay geometrically while  $\rho$  is preserved.*

Different values of  $\rho$  define qualitatively different economic regimes:

Regime	$\rho$ range	Equilibrium distribution	Economic character
Strong complements	$\rho < 0$	Power-law ( $q$ -exponential, $q < 1$ )	Bottleneck-dominated
Unit elasticity	$\rho = 0$	Log-normal (Gibbs-Boltzmann)	Balanced
Weak substitutes	$0 < \rho < 1$	Stretched exponential ( $q > 1$ )	Abundance-dominated
Perfect substitutes	$\rho = 1$	Degenerate (winner-take-all)	Commodity

The critical boundary  $\rho = 0$  (Cobb-Douglas) separates complementary from substitutable regimes. At this boundary, the equilibrium distribution is exactly log-normal—the standard Gibbs-Boltzmann distribution of statistical mechanics. Moving away from  $\rho = 0$  in either direction “deforms” the distribution, generating heavier or lighter tails.

## 7.2 Connection to Tsallis Thermodynamics

The identification  $q = \rho$  connects the CES framework to Tsallis non-extensive statistical mechanics (Tsallis, 1988). Tsallis introduced the deformation parameter  $q$  to handle systems with long-range correlations, where the standard Boltzmann-Gibbs framework (which assumes weak correlations and extensivity) breaks down.

In the economic context, the mapping is:

- $q = 1$  ( $\rho = 1$ , perfect substitutes): agents are independent, entropy is extensive (Shannon), distributions are exponential. This is the standard competitive equilibrium.
- $q < 1$  ( $\rho < 1$ , complements): agents are *positively correlated* through the complementarity of the production function. Entropy is sub-extensive. Distributions are heavy-tailed. Small perturbations to bottleneck inputs have outsized effects on the aggregate.
- $q > 1$ : not economically relevant for production (would imply  $\rho > 1$ , an increasing-returns CES), but appears in consumption aggregation where goods can be “super-substitutes.”

**Proposition 7.2** (Tail exponent). *The equilibrium allocation under  $F_q = \Phi_{CES}(q) - T \cdot S_q$  has tail behavior*

$$p_j \sim \varepsilon_j^{-1/(1-q)} \quad \text{as } \varepsilon_j \rightarrow \infty, \tag{13}$$

for  $q < 1$ . The Pareto exponent is  $\zeta = 1/(1 - \rho)$ , which is the elasticity of substitution  $\sigma$ .

This provides a deep explanation for the ubiquity of power laws in economics (Gabaix, 2009): they emerge from complementary production ( $\rho < 0$ ) combined with entropy maximization at  $T > 0$  (Remark 5.3). The Pareto exponent is not a free parameter—it is determined by the production technology through  $\sigma = 1/(1 - \rho)$ . More precisely,  $\rho$

determines the *shape* of the tail (the exponent  $\zeta$ ), while  $T$  determines whether the tail *exists at all* (it vanishes as  $T \rightarrow 0$ ).

### 7.3 The $\rho$ Phase Diagram

Combining the universality class with the temperature  $T$  yields the complete phase diagram of economic aggregation:

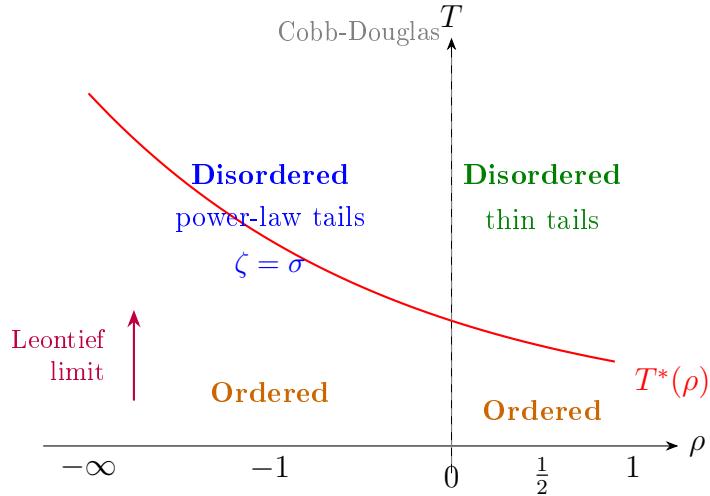


Figure 2: Phase diagram of economic aggregation in  $(\rho, T)$  space. The critical curve  $T^*(\rho)$  separates ordered (low- $T$ , efficient) from disordered (high- $T$ , noisy) regimes. Below the critical curve, allocations are approximately optimal; above it, entropy dominates. The Cobb-Douglas boundary ( $\rho = 0$ ) separates power-law equilibria (left) from thin-tailed equilibria (right). The tail exponent  $\zeta = \sigma = 1/(1 - \rho)$  is determined by the universality class.

## 8 Implications for the $(\rho, T)$ Framework

### 8.1 Axiom Reduction

The  $(\rho, T)$  economic free energy framework (Smirl, 2026a) rests on five axioms. Axiom A1 states: “Production at each level is CES.” Theorem 1.1 shows that A1 can be replaced by two weaker axioms:

Original	Replacement	Content
A1: CES production	A1a: Constant returns	Homogeneity of degree 1
	A1b: Scale consistency	Associative aggregation

This is not merely an aesthetic improvement. The original A1 is a parametric assumption about functional form—the kind of assumption that invites the response “but what

if production isn't CES?" The replacements are structural properties: constant returns is standard in aggregation theory, and scale consistency is the requirement that macro quantities are well-defined (independent of arbitrary partitioning choices). The response "what if aggregation isn't scale-consistent?" is much harder to sustain—it amounts to claiming that economic aggregates depend on the analyst's choice of sector boundaries.

## 8.2 Why $\rho$ Is the Natural Parameter

The emergence result explains why  $\rho$ —rather than  $\sigma = 1/(1-\rho)$ , or  $K = (1-\rho)(J-1)/J$ , or any other reparameterization—is the natural parameter of the framework.

1. **RG invariance:**  $\rho$  is the exponent preserved under coarse-graining.  $\sigma$  and  $K$  are derived quantities.
2. **Entropy matching:**  $\rho$  equals the Rényi order  $\alpha$ . No other parameterization has this property.
3. **Universality class label:**  $\rho$  determines the equilibrium distribution's tail behavior.  $\sigma$  does too (it's a function of  $\rho$ ), but  $\rho$  is the primitive object—the power in the power mean.
4. **Tsallis deformation:**  $\rho = q$  in the Tsallis framework. The deformation of thermodynamics is parameterized by  $\rho$  directly.

That said, different derived quantities are natural in different contexts. The curvature  $K = (1-\rho)(J-1)/J$  is natural for the quadruple-role theorem because it measures the gap between CES and linear aggregation at the symmetric point. The elasticity  $\sigma = 1/(1-\rho)$  is natural for demand analysis because it directly gives the percentage response to relative price changes. The emergence result says that  $\rho$  is the *fundamental* parameter from which these are derived.

## 8.3 The Translog as an Irrelevant Perturbation

The translog production function ([Christensen, Jorgenson, and Lau, 1973](#)) is the standard "flexible" alternative to CES:

$$\log F(\mathbf{x}) = \alpha_0 + \sum_j \alpha_j \log x_j + \frac{1}{2} \sum_{j,k} \beta_{jk} \log x_j \log x_k. \quad (14)$$

The translog nests Cobb-Douglas ( $\beta_{jk} = 0$ ) but not general CES. From the RG perspective, the translog adds irrelevant operators to the Cobb-Douglas fixed point: the  $\beta_{jk}$  terms break scale consistency and vanish under coarse-graining. The translog is not an alternative to CES—it is a transient deviation from CES that exists only at fine scales.

This resolves a long-standing puzzle in production economics: why CES fits aggregate data well despite being “more restrictive” than translog. The answer is that the translog’s extra flexibility is precisely the kind of flexibility that doesn’t survive aggregation. Estimating a translog at the macro level is fitting noise—the irrelevant operators that haven’t yet been averaged away.

## 8.4 Weighted CES and Asymmetric Inputs

The emergence theorem (Theorem 1.1) assumes symmetric inputs. Real economies have heterogeneous input shares. The weighted CES

$$F_\rho(\mathbf{x}; \mathbf{w}) = \left( \sum_{j=1}^J w_j x_j^\rho \right)^{1/\rho}, \quad w_j > 0, \quad \sum_j w_j = 1, \quad (15)$$

is *not* a fixed point of the symmetric RG—the weights break the symmetry. However, weighted CES is the fixed point of the *weighted* aggregation RG, where blocks have unequal sizes and the RG preserves the relative weights. The emergence result generalizes:

**Corollary 8.1** (Weighted emergence). *Let  $F : \mathbb{R}_+^J \rightarrow \mathbb{R}_+$  be continuous, strictly increasing, homogeneous of degree one, and scale-consistent under the weighted partition  $\{B_1, \dots, B_m\}$  with  $w_i = |B_i|/J$ . Then  $F$  is weighted CES with weights  $w_j$  proportional to the atoms of the partition.*

The weights  $w_j$  are *not* preserved under coarse-graining in the same way as  $\rho$ : they can shift as the partition changes. In RG language, the weights are “dangerously irrelevant”—irrelevant in the scaling sense but capable of affecting non-universal amplitudes. Economically, this means the input shares can vary across levels of aggregation (sector shares  $\neq$  firm shares), but the elasticity of substitution is the same at every level.

# 9 Empirical Predictions

The emergence of CES generates five testable predictions that distinguish it from the alternative hypothesis that CES is merely a convenient approximation.

## 9.1 Prediction 1: CES Fit Improves with Aggregation

If CES is the RG fixed point and non-CES deviations are irrelevant operators, then CES should fit progressively better at higher levels of aggregation. Specifically:

**Prediction 9.1.** Let  $R_\ell^2$  denote the  $R^2$  of the Kmenta approximation (second-order Taylor expansion of CES) relative to the exact production function at aggregation level  $\ell$  (establishment  $\rightarrow$  firm  $\rightarrow$  industry  $\rightarrow$  sector  $\rightarrow$  economy). Then  $R_{\ell+1}^2 > R_\ell^2$ .

**Data:** US Census of Manufactures (establishment-level) aggregated to progressively coarser industry classifications (6-digit  $\rightarrow$  4-digit  $\rightarrow$  3-digit  $\rightarrow$  2-digit NAICS).

**Test:** Estimate translog at each level; test whether the  $\beta_{jk}$  coefficients shrink toward CES-consistency ( $\beta_{jk} \rightarrow 0$  for  $j \neq k$ ,  $\beta_{jj} \rightarrow$  common value).

## 9.2 Prediction 2: $\rho$ Is Preserved Across Scales

**Prediction 9.2.** The elasticity of substitution  $\sigma$  estimated at different aggregation levels should be approximately constant:  $\hat{\sigma}_\ell \approx \hat{\sigma}_{\ell'}$  for all  $\ell, \ell'$ .

This is sharper than Prediction 1. It says not only that CES fits better at higher aggregation, but that the *same*  $\rho$  fits at every level. Violations would indicate that the RG flow has not converged, which is informative about the “RG distance” between the current scale and the fixed point.

**Data:** Same as Prediction 1, plus cross-country comparisons using Penn World Table capital-labor substitution estimates at different levels of industry aggregation.

## 9.3 Prediction 3: Tail Exponents Match $\sigma$

**Prediction 9.3.** In sectors with estimated elasticity of substitution  $\hat{\sigma}$ , the Pareto tail exponent of the firm size distribution should satisfy  $\hat{\zeta} \approx \hat{\sigma}$ .

This prediction connects the production side ( $\rho$ ) to the distributional side (tail exponent) through the Tsallis equilibrium. It is falsifiable: if firm sizes follow a power law for reasons unrelated to the production technology (e.g., preferential attachment in firm growth), then  $\hat{\zeta}$  and  $\hat{\sigma}$  will be unrelated.

**Data:** Bureau of Labor Statistics firm size distributions by 3-digit NAICS, combined with sector-level  $\sigma$  estimates from [Oberfield and Raval \(2014\)](#).

## 9.4 Prediction 4: Translog Coefficients Shrink Under Aggregation

**Prediction 9.4.** In a translog regression  $\log F = \alpha_0 + \sum \alpha_j \log x_j + \frac{1}{2} \sum \beta_{jk} \log x_j \log x_k + \varepsilon$ , the ratio  $\|\boldsymbol{\beta}\|_F / \|\boldsymbol{\alpha}\|$  (Frobenius norm of the interaction matrix relative to the first-order coefficients) decreases with aggregation level.

This is the direct test of irrelevant operators shrinking under the RG. The translog interaction terms  $\beta_{jk}$  are the leading irrelevant operators; they should decrease as  $O(1/k^\ell)$  where  $k$  is the block size and  $\ell$  is the number of RG steps.

## 9.5 Prediction 5: Complementary Sectors Have Heavier Tails

**Prediction 9.5.** Cross-sectionally, sectors with lower estimated  $\rho$  (higher complementarity) should exhibit heavier-tailed firm size distributions:  $\partial\hat{\zeta}/\partial\hat{\rho} > 0$ , provided  $T > 0$  (information friction is non-negligible).

This combines the universality class result (Proposition 7.2) with cross-sectional variation. High-complementarity sectors (construction, healthcare, complex manufacturing) should have more extreme firm-size inequality than low-complementarity sectors (commodity agriculture, basic materials). The qualification  $T > 0$  is essential (Remark 5.3): in a hypothetical perfectly optimized sector ( $T \rightarrow 0$ ), complementarity compresses rather than stretches the distribution. In practice, all real sectors have  $T > 0$ , so the prediction applies; but heavily planned or regulated sectors (lower effective  $T$ ) may show weaker tail effects than market-driven ones at the same  $\rho$ .

# 10 Extensions

## 10.1 Nested CES and the RG Flow

The hierarchical CES structure used in the  $(\rho, T)$  framework—different  $\rho$  at each level—corresponds to a *non-trivial RG trajectory*: the system has not converged to a single fixed point but instead traces a path through the space of  $\rho$  values as the scale changes.

**Definition 10.1** (RG  $\beta$ -function for  $\rho$ ). The RG flow of  $\rho$  under coarse-graining from scale  $s$  to scale  $s + ds$  is

$$\frac{d\rho}{d\log s} = \beta(\rho), \quad (16)$$

where  $\beta$  encodes how the effective complementarity changes with scale.

For the CES fixed points,  $\beta(\rho) = 0$ — $\rho$  is exactly preserved. But if the micro-level production is not exactly CES, or if different scales have genuinely different production structures (e.g., strong complementarity within firms but substitutability across firms), then  $\beta(\rho) \neq 0$  and the effective  $\rho$  flows.

The fixed points of  $\beta$  are the candidates for “natural” values of  $\rho$  at the macro level:

- $\rho = 1$  (perfect substitutes): always a fixed point. Stable from above (no  $\rho > 1$  for production).
- $\rho = 0$  (Cobb-Douglas): a fixed point if the micro distribution of productivities is log-normal (by CLT). Stable under mild conditions.
- $\rho \rightarrow -\infty$  (Leontief): a fixed point. Unstable—any substitutability, however small, pulls  $\rho$  away from  $-\infty$  under aggregation.

The empirical elasticity of substitution for aggregate production ( $\sigma \approx 0.6\text{--}0.8$  in most estimates, implying  $\rho \approx -0.25$  to  $-0.67$ ) sits between Cobb-Douglas and Leontief, suggesting that the macro economy has not converged to either fixed point. This is consistent with the “slow RG” hypothesis: the economy is still flowing toward  $\rho = 0$ , but the flow is slow enough that the observed  $\rho$  remains negative.

## 10.2 CES Emergence and Paper 15 (Endogenous $\rho$ )

Paper 15 in the thesis series (Smirl, 2026b) shows that  $\rho$  is endogenous: it evolves through firm optimization, evolutionary selection (the Price equation), and technological standardization (Wright’s Law). The emergence result adds a fourth channel: *aggregation itself* can change the effective  $\rho$  through the RG flow.

These four channels operate at different scales:

Channel	Scale	Direction
Firm optimization	Within firm	Procyclical ( $\rho$ rises in booms)
Evolutionary selection	Across firms	Diversity-preserving (Price equation)
Wright’s Law standardization	Technology	Secular increase ( $\rho$ rises over time)
Aggregation RG	Cross-scale	Flow toward Cobb-Douglas ( $\rho \rightarrow 0$ )

The tension between within-scale dynamics (which can push  $\rho$  in either direction) and cross-scale aggregation (which pulls toward  $\rho = 0$ ) creates a rich dynamical system whose equilibrium  $\rho$  depends on the relative strength of the channels.

## 10.3 Beyond Symmetric CES

The emergence theorem assumes symmetry (identical treatment of all inputs). Real production functions are asymmetric: capital and labor have different shares, skilled and unskilled workers are not interchangeable, etc. How much of the emergence result survives?

The key distinction is between *share asymmetry* (different  $w_j$ ) and *elasticity asymmetry* (different  $\rho$  for different input pairs). Corollary 8.1 shows that share asymmetry is compatible with emergence: weighted CES is the fixed point of the weighted RG. But elasticity asymmetry—different  $\rho_{jk}$  for different pairs  $(j, k)$ —breaks the emergence result. There is no simple fixed point when substitution elasticities vary across input pairs.

However, the RG still operates. Under coarse-graining, the effective elasticity matrix  $\rho_{jk}$  is averaged across inputs within each block. If the within-block variation in  $\rho_{jk}$  is small relative to the between-block variation, the block-level aggregation is approximately CES with  $\rho$  equal to the average within-block elasticity. This is a *local* emergence: CES emerges within homogeneous blocks even when the global structure is non-CES.

## 11 Conclusion

The CES production function is not an assumption. It is a theorem—the unique consequence of constant returns to scale and scale consistency. Three independent lines of argument (renormalization group, functional equations, maximum entropy self-consistency) converge on this conclusion, and their convergence is itself a mathematical fact (Theorem 6.1).

The implications ramify through the  $(\rho, T)$  framework:

1. The foundational axiom “production is CES” is replaced by two structural properties that any multi-scale economy must satisfy. The framework rests on firmer ground.
2. The parameter  $\rho$  is identified as a universality class label—the one feature of the production technology that survives aggregation. This explains why CES works empirically: every other feature of the production function is irrelevant in the RG sense.
3. The matching  $\rho = \alpha$  (CES parameter = Rényi order) locks the production technology to the entropy measure, generating the Tsallis free energy as the natural thermodynamic potential. Heavy-tailed distributions emerge endogenously from complementary production.
4. The translog and other flexible functional forms are reinterpreted as irrelevant perturbations of the CES fixed point—useful for detecting non-CES features at fine scales, but guaranteed to vanish under aggregation.
5. Five empirical predictions (Section 9) distinguish the emergence hypothesis from the alternative that CES is merely a convenient approximation.

In the language of statistical mechanics: CES is to production economics what the Gaussian is to probability theory. Not chosen for convenience. Not assumed for tractability. Inevitable.

## References

- Aczél, János. 1948. “On Mean Values.” *Bulletin of the American Mathematical Society* 54(4): 392–400.
- Arrow, Kenneth J., Hollis B. Chenery, Bagicha S. Minhas, and Robert M. Solow. 1961. “Capital-Labor Substitution and Economic Efficiency.” *Review of Economics and Statistics* 43(3): 225–250.
- Christensen, Laurits R., Dale W. Jorgenson, and Lawrence J. Lau. 1973. “Transcendental Logarithmic Production Frontiers.” *Review of Economics and Statistics* 55(1): 28–45.
- Gabaix, Xavier. 2009. “Power Laws in Economics and Finance.” *Annual Review of Economics* 1(1): 255–294.
- Hardy, G. H., J. E. Littlewood, and G. Pólya. 1952. *Inequalities*. 2nd ed. Cambridge: Cambridge University Press.
- Kolmogorov, Andrey N. 1930. “Sur la Notion de la Moyenne.” *Atti della Reale Accademia Nazionale dei Lincei. Rendiconti* 12: 388–391.
- Nagumo, Mitio. 1930. “Über eine Klasse der Mittelwerte.” *Japanese Journal of Mathematics* 7: 71–79.
- Oberfield, Ezra, and Devesh Raval. 2014. “Micro Data and Macro Technology.” NBER Working Paper No. 20452.
- Sato, Kazuo. 1967. “A Two-Level Constant-Elasticity-of-Substitution Production Function.” *Review of Economic Studies* 34(2): 201–218.
- Smirl, Jon. 2026a. “The  $(\rho, T)$  Framework: Five Axioms, One Phase Diagram, and the Deductive Structure of Economic Free Energy.” Working Paper.
- Smirl, Jon. 2026b. “Endogenous Complementarity: The Self-Referential Dynamics of  $\rho$ .” Working Paper.
- Tsallis, Constantino. 1988. “Possible Generalization of Boltzmann-Gibbs Statistics.” *Journal of Statistical Physics* 52(1–2): 479–487.