

Emergent CES and the Quadruple Role of Curvature

Why Constant Elasticity of Substitution Is Not an Assumption,
and What the Forced Form Does

Jon Smirl

Independent Researcher

February 2026

Working Paper

Abstract

CES is not an assumption—it is forced by economic structure. [Houthakker \(1955\)](#) showed one special case: Leontief firms with Pareto-distributed coefficients aggregate to Cobb-Douglas. This paper proves the general result: *any* continuous, symmetric, constant-returns production function that is self-consistent across aggregation levels must be CES, for some ρ . The proof proceeds via three independent arguments—aggregation fixed points, Aczél’s functional equation characterization, and maximum entropy self-consistency—shown to be three projections of a single geometric fact. A stability theorem establishes that even when micro-level production is *not* CES, repeated aggregation drives the macro function toward CES at geometric rate $O(k^{-L/2})$, with the curvature parameter ρ preserved exactly and all other functional-form details washed out. The parameter ρ thus emerges as an aggregation-invariant class label—the one feature of production technology that survives multi-scale aggregation—locked to Rényi entropy of matching order ($q = \rho$). Part II proves that the forced functional form does four things simultaneously, all controlled by a single dimensionless parameter $K = (1 - \rho)(J - 1)/J$: (a) superadditivity gap = $\Omega(K) \cdot \text{diversity}$; (b) diversity-encoding sensitivity = $\Theta(K) \cdot \text{idiosyncratic dispersion}$; (c) strategic manipulation penalty = $-\Omega(K) \cdot \text{deviation}^2$; (d) network scaling exponent = $1/\rho$. Properties (a)–(c) are three views of the curvature of CES isoquants at the cost-minimizing point. Empirical tests on NBER-CES manufacturing data confirm the key predictions: translog interaction terms decay monotonically under NAICS aggregation, and CES fit improves from $R^2 = 0.94$ at 6-digit to $R^2 = 0.98$ at 2-digit.

Keywords: CES production function, multi-scale aggregation, power means, Rényi entropy, isoquant curvature, superadditivity, diversification, strategic independence, network scaling, secular equation

JEL: C60, C62, D20, D24, D43, D81, E10, L13

Contents

1	Introduction	4
2	Preliminaries: CES, Power Means, and Entropy	6
2.1	The CES Aggregate	6
2.2	The Curvature Parameter	7
2.3	Quasi-Arithmetic Means	7
2.4	Rényi and Tsallis Entropies	7
3	Argument 1: Multi-Scale Aggregation Fixed Point	8
3.1	The Aggregation Operator	8
3.2	CES as the Unique Fixed Point	8
3.3	Stability: What the Aggregation Framework Adds	9
3.4	The ρ Eigenvalue	15
4	Argument 2: Functional Equations	16
4.1	Aczél’s Characterization	16
4.2	What the Functional Equation Argument Adds	17
5	Argument 3: Maximum Entropy Self-Consistency	17
5.1	The Entropy-Allocation Loop	17
5.2	The Matching of Exponents	20
5.3	The Economic Potential Decomposition	21
6	Convergence of the Three Arguments	21
7	The ρ Aggregation-Invariant Class	23
7.1	ρ as an Aggregation-Invariant Exponent	23
7.2	Connection to Tsallis Non-Extensive Economics	23
7.3	The (ρ, T) Regime Diagram	24
7.4	Implications for the Framework	24
7.5	Normalization and Cross-Economy Comparison	25

8	The Curvature Lemma	26
8.1	The Symmetric Point	26
8.2	Gradient and Hessian at the Symmetric Point	26
8.3	Isoquant Curvature	26
9	Superadditivity	27
10	Correlation Robustness	28
10.1	Setup	28
10.2	The Theorem	28
11	Strategic Independence	30
11.1	Setup	30
11.2	The Theorem	30
12	Network Scaling	31
13	The Unified Theorem	31
13.1	The Geometric Intuition	32
13.2	Unified Scaling: All Three Are $\Theta(K)$	33
14	General Weights and the Secular Equation	33
14.1	Effective Shares and the Cost-Minimizing Point	33
14.2	The Hessian at General Weights	33
14.3	The Secular Equation	33
14.4	The Generalized Curvature Parameter	34
14.5	General-Weight Quadruple Role	34
15	Empirical Predictions	34
15.1	Monte Carlo: CES Emerges from Non-CES Micro Data	35
15.2	NBER-CES Manufacturing Database	35
15.3	FRED Industrial Production Hierarchy	36
16	Extensions and Discussion	36
16.1	Nested CES and Multi-Scale Structure	36
16.2	Weighted CES and Asymmetric Inputs	38
16.3	The Translog as an Irrelevant Perturbation	38
16.4	Tightness of the Quadruple Role Bounds	38
16.5	What This Paper Does Not Cover	39

Part I: Emergence — CES Is Forced

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 a *theorem*. Just as the Central Limit Theorem shows that the Gaussian is forced by repeated summation of independent random variables regardless of their individual distributions, CES is forced by repeated aggregation of production units regardless of their individual technologies.

The idea that aggregation can produce CES has a distinguished precursor. [Houthakker \(1955\)](#) showed that if firms use Leontief (fixed-coefficient) technology and their capital-labor ratios follow a Pareto distribution, the aggregate production function is Cobb-Douglas. Subsequent work extended this substantially. [Jones \(2005\)](#) showed that CES aggregates arise from Cobb-Douglas firms with Fréchet-distributed productivities, for specific values of ρ determined by the tail index. [Oberfield and Raval \(2014\)](#) took the most ambitious empirical approach: using plant-level Census data, they estimated micro-level elasticities of substitution and showed that these are systematically lower than macro estimates, with the gap explained by between-plant reallocation—precisely the aggregation channel this paper formalizes. But each of these results relies on specific distributional or functional-form assumptions at the micro level. The general question—*what structural properties of an economy force CES at the aggregate level, without any distributional assumption on firm-level technology?*—has remained open.

This paper answers that question. Two structural properties suffice:

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 grouping: 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.

Axiom 2 connects to a long tradition in aggregation theory. Gorman (1968) posed the “consistent aggregation” problem: under what conditions does a macro production function faithfully represent underlying micro technologies? Blackorby and Schworm (1993) formalized this as a functional equation constraint and showed that consistent aggregation imposes severe restrictions on the class of admissible functions. The scale consistency axiom is the natural strengthening of their requirement to *multi-level* hierarchies: not just “micro aggregates consistently to macro,” but “every intermediate level aggregates consistently to every higher level.” As Theorem 1.1 shows, this multi-level requirement is what selects CES from the larger class of consistently aggregable functions.

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).

No assumption is made about the distribution of firm-level productivities, the network structure of intermediate inputs, or the functional form of micro-level technology. CES is not selected from a menu of candidate functional forms—it is the only possibility.

Moreover, a stability theorem (Theorem 3.4) shows that CES is not merely the unique exact solution but also an *attractor*: even when micro-level production violates scale consistency, repeated aggregation drives the macro function toward CES at geometric rate $O(k^{-L/2})$, preserving the curvature parameter ρ exactly while eliminating all other functional-form detail. This is the analogue of the Central Limit Theorem for production aggregation: CES plays the role of the Gaussian, ρ plays the role of the variance, and non-CES deviations play the role of higher cumulants that vanish under summation.

Unlike Houthakker’s result, the stability theorem requires no distributional assumption and yields the full CES family, not just Cobb-Douglas.

Part I develops three independent proofs of Theorem 1.1 (Sections 3 to 5), shows their convergence (Section 6), and identifies ρ as an aggregation-invariant class label (Section 7). Part II proves that the forced functional form has four simultaneous consequences controlled by ρ —three curvature effects governed by $K = (1 - \rho)(J - 1)/J$ and one algebraic scaling law governed by ρ directly (Section 8–Section 13)—and extends to general weights (Section 14).

2. Preliminaries: CES, Power Means, and Entropy

This section establishes the canonical definitions used throughout the paper.

2.1 The CES Aggregate

For $J \geq 2$ components, the **CES aggregate** with weights $a_j > 0$ summing to 1 is

$$F(\mathbf{x}) = \left(\sum_{j=1}^J a_j x_j^\rho \right)^{1/\rho}, \quad \mathbf{x} = (x_1, \dots, x_J) \in \mathbb{R}_+^J \quad (1)$$

where $\rho < 1$, $\rho \neq 0$, is the **substitution parameter**. The **elasticity of substitution** is $\sigma = 1/(1 - \rho)$.

The components are called *complements* when $\rho < 0$ ($\sigma < 1$) and *weak complements* when $0 < \rho < 1$ ($\sigma > 1$ but finite). The boundary cases are:

$$\rho = 1 : \quad F(\mathbf{x}) = \sum_{j=1}^J a_j x_j \quad (\text{linear: perfect substitutes}), \quad (2)$$

$$\rho \rightarrow 0 : \quad F(\mathbf{x}) = \prod_{j=1}^J x_j^{a_j} \quad (\text{Cobb-Douglas}), \quad (3)$$

$$\rho \rightarrow -\infty : \quad F(\mathbf{x}) = \min_j x_j \quad (\text{Leontief: perfect complements}). \quad (4)$$

With equal weights $a_j = 1/J$, the CES aggregate is precisely the **power mean** $M_\rho(\mathbf{x})$ of order ρ :

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

Remark 2.1. F is concave and homogeneous of degree 1 for all $\rho < 1$. Both properties are standard and are used freely throughout.

2.2 The Curvature Parameter

Definition 2.2 (Curvature parameter). The **curvature parameter** of the equal-weight CES aggregate with J components is

$$K = (1 - \rho) \frac{J - 1}{J}. \quad (6)$$

Properties: (i) $K > 0$ for all $\rho < 1$. (ii) K is strictly increasing in $(1 - \rho)$ and in J . (iii) $K \rightarrow \infty$ as $\rho \rightarrow -\infty$ (Leontief limit). (iv) $K \rightarrow 0$ as $\rho \rightarrow 1^-$ (perfect substitutes). (v) At Cobb-Douglas ($\rho = 0$): $K = (J - 1)/J$.

2.3 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

$$M_\varphi(\mathbf{x}) = \varphi^{-1} \left(\frac{1}{J} \sum_{j=1}^J \varphi(x_j) \right). \quad (7)$$

Power means correspond to $\varphi(x) = x^\rho$ (with $\varphi(x) = \log x$ for $\rho = 0$). The quasi-arithmetic family is strictly larger. The question is: what selects power means from the larger quasi-arithmetic family?

2.4 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

$$H_\alpha(\mathbf{p}) = \frac{1}{1 - \alpha} \log \left(\sum_{j=1}^J p_j^\alpha \right). \quad (8)$$

The $\alpha \rightarrow 1$ limit is Shannon entropy. The *Tsallis entropy* of order q is

$$S_q(\mathbf{p}) = \frac{1}{q - 1} \left(1 - \sum_{j=1}^J p_j^q \right). \quad (9)$$

The sufficient statistic for both is $\sum_j p_j^\alpha$ —a power sum, the same object that appears inside CES.

3. Argument 1: Multi-Scale Aggregation Fixed Point

3.1 The Aggregation Operator

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. After L iterations, a single aggregate output remains.

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

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

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 operator if $\mathcal{R}_k f^* = f^*$ for all block sizes k —meaning the functional form is preserved under aggregation.

3.2 CES as the Unique Fixed Point

Theorem 3.2 (Aggregation fixed point). *Among continuous, symmetric, strictly increasing, homogeneous-of-degree-one functions $f : \mathbb{R}_+^k \rightarrow \mathbb{R}_+$, the power means M_ρ 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 :

$$f(x_1, \dots, x_{mk}) = f(f(x_1, \dots, x_k), f(x_{k+1}, \dots, x_{2k}), \dots, f(x_{(m-1)k+1}, \dots, x_{mk})). \quad (10)$$

Setting φ = the quasi-arithmetic generator of f (which exists by the Kolmogorov-Nagumo theorem since f is a continuous symmetric mean), (10) 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 φ 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 φ , 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 $\lambda > 0$. This is a functional equation in φ . 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_ρ . \square

Remark 3.3 (The two-step structure). The proof reveals that uniqueness proceeds in two logically distinct steps, and it is important to be clear about what each contributes:

- (i) **Associativity** \rightarrow **quasi-arithmetic**. Scale consistency is precisely the associativity condition of the Kolmogorov-Nagumo theorem (Theorem 4.2), which restricts the function to the quasi-arithmetic family. But as the proof shows, associativity is *automatic* for any quasi-arithmetic mean—the nesting condition imposes no additional constraint within this family. The aggregation operator \mathcal{R}_k has a very large fixed-point set: all quasi-arithmetic means, including $\varphi(x) = e^x$, $\varphi(x) = \log x$, etc.
- (ii) **Homogeneity** \rightarrow **power means**. It is Aczél’s characterization (Theorem 4.3)—not the aggregation operator—that selects power means from the quasi-arithmetic family. Homogeneity of degree one forces $\varphi(x) = cx^\rho$.

Taken alone, Theorem 3.2 is therefore a repackaging of classical results: the aggregation “fixed point” framing does not add a new restriction beyond associativity plus homogeneity. Its value lies elsewhere—in the *stability* result that follows.

3.3 Stability: What the Aggregation Framework Adds

Theorem 3.2 characterizes fixed points within the class of functions that already satisfy scale consistency. But real economies need not be scale-consistent at any finite level. Firm-level production may involve threshold effects, capacity constraints, or other non-quasi-arithmetic features. The question is: does CES *emerge* from repeated aggregation even when the micro-level function is not CES?

The idea that CES can emerge from micro-level heterogeneity has a distinguished precursor. [Houthakker \(1955\)](#) showed that if firm-level technologies are Leontief (fixed coefficients) but the coefficient ratios follow a Pareto distribution across firms, the aggregate production function is Cobb-Douglas. Houthakker’s result is a special case ($\rho = 0$) of a more general phenomenon: aggregation over heterogeneous micro-level technologies drives the macro production function toward CES. Theorem 3.4 below formalizes this for arbitrary ρ and without distributional assumptions on the micro technology.

This is where the aggregation framework contributes something the classical characterization theorems do not. The aggregation operator \mathcal{R}_k acts on the *full* space \mathcal{F} of continuous, symmetric, strictly increasing, homogeneous-of-degree-one functions—not just on quasi-arithmetic means. A function $f \in \mathcal{F}$ that is not quasi-arithmetic violates associativity: its aggregate depends on how you draw the block boundaries. Under \mathcal{R}_k , this boundary dependence shrinks.

Theorem 3.4 (Attraction to CES). *Let $f \in \mathcal{F}$ be continuous, symmetric, strictly increasing, and homogeneous of degree one, with second-order behavior near the symmetric point characterized by $\rho_f = \lim_{\gamma \rightarrow 0} [\log f(\mathbf{x}) - \log \bar{x}] / [\gamma^2/2]$ (the Kmenta curvature, after [Kmenta 1967](#)). Then:*

(i) *After L applications of \mathcal{R}_k , the distance from the power mean M_{ρ_f} satisfies*

$$\|\mathcal{R}_k^L f - M_{\rho_f}\|_{\infty, B_\gamma} = O(k^{-L/2}) \cdot \|f - M_{\rho_f}\|_{\infty, B_\gamma} \quad (11)$$

where $B_\gamma = \{\mathbf{x} : \|\mathbf{x}/\bar{x} - \mathbf{1}\|_\infty < \gamma\}$ is a neighborhood of the symmetric point and $\gamma > 0$ is small enough that the Taylor expansion converges.

(ii) *The Kmenta curvature ρ_f is preserved exactly: $\rho_{\mathcal{R}_k f} = \rho_f$.*

(iii) *The m -th order deviation from CES contracts at rate $O(k^{-(m/2-1)})$ per step; the slowest mode ($m = 3$) contracts at $O(k^{-1/2})$.*

Proof. The proof proceeds by expressing f and M_{ρ_f} in a common basis (the power sums s_m), showing that the aggregation operator preserves the s_2 coefficient while contracting every s_m coefficient for $m \geq 3$, then iterating.

Step 1 (Power-sum expansion). Write $x_j = \bar{x}(1 + \delta_j)$ where $\bar{x} = \frac{1}{n} \sum x_j$ and $\sum \delta_j = 0$. By homogeneity of degree one, $f(\mathbf{x}) = \bar{x} \cdot f(\mathbf{1} + \boldsymbol{\delta})$, so it suffices to expand $f(\mathbf{1} + \boldsymbol{\delta})$ in powers of $\boldsymbol{\delta}$. By symmetry, f depends on $\boldsymbol{\delta}$ only through the symmetric polynomials. Because $\sum \delta_j = 0$, the independent symmetric functions are the power sums $s_m := \frac{1}{n} \sum_{j=1}^n \delta_j^m$ for $m \geq 2$. Provided $\|\boldsymbol{\delta}\|_\infty < \gamma$ with γ small enough that f is analytic on B_γ , there is a convergent expansion

$$f(\mathbf{x}) = \bar{x} \left(1 + c_2 s_2 + \sum_{m=3}^M a_m s_m + E_M(\boldsymbol{\delta}) \right), \quad (12)$$

where $|E_M(\boldsymbol{\delta})| \leq C_M \gamma^{M+1}$ for a constant C_M depending on f . Here $c_2 = (\rho_f - 1)/2$ defines the Kmenta curvature ρ_f —so named because [Kmenta \(1967\)](#) showed that $\log M_\rho(\mathbf{x}) \approx \log \bar{x} + (\rho - 1)s_2/2$ is the second-order approximation to CES—and $s_2 = \text{Var}(\mathbf{x})/\bar{x}^2$. For the power mean M_{ρ_f} , the expansion (12) has $a_m = 0$ for all $m \geq 3$ (which can be verified by direct Taylor expansion of M_ρ).

Step 2 (ANOVA decomposition of one aggregation step). Under \mathcal{R}_k , the $n = k \cdot r$ inputs are partitioned into r blocks of size k . Decompose the deviation of the j -th element of block i as

$$\delta_{ij} = \bar{\delta}_i + \epsilon_{ij}, \quad \bar{\delta}_i := \frac{1}{k} \sum_{j=1}^k \delta_{ij}, \quad \sum_{j=1}^k \epsilon_{ij} = 0.$$

Here $\bar{\delta}_i$ is the block- i mean deviation and ϵ_{ij} is the within-block residual. Define the within-block power sums $s_m^{(i)} := \frac{1}{k} \sum_{j=1}^k \delta_{ij}^m$. By (12), the block output is

$$Y_i = f(\mathbf{x}_i) = \bar{x} \left(1 + c_2 s_2^{(i)} + \sum_{m=3}^M a_m s_m^{(i)} + E_M^{(i)} \right), \quad (13)$$

where $|E_M^{(i)}| \leq C_M \gamma^{M+1}$.

Step 3 (Binomial decomposition of within-block power sums). Expanding $\delta_{ij}^m = (\bar{\delta}_i + \epsilon_{ij})^m$ by the binomial theorem and averaging over j :

$$s_m^{(i)} = \bar{\delta}_i^m + \underbrace{\binom{m}{1} \bar{\delta}_i^{m-1} \frac{1}{k} \sum_j \epsilon_{ij}}_{=0} + \binom{m}{2} \bar{\delta}_i^{m-2} \frac{1}{k} \sum_j \epsilon_{ij}^2 + R_m^{(i)}, \quad (14)$$

where $R_m^{(i)}$ collects terms involving $\frac{1}{k} \sum_j \epsilon_{ij}^p$ with $p \geq 3$, or mixed products of total degree ≥ 3 in ϵ . The $\binom{m}{1}$ term vanishes by $\sum_j \epsilon_{ij} = 0$.

For $m = 2$ specifically:

$$s_2^{(i)} = \bar{\delta}_i^2 + \frac{1}{k} \sum_j \epsilon_{ij}^2. \quad (15)$$

This is the one-way ANOVA identity: within-block second moment = squared block mean + within-block variance.

Step 4 (Preservation of ρ —the second-order coefficient). The aggregated function $\mathcal{R}_k f$ takes the block outputs (Y_1, \dots, Y_r) as inputs. Define $\bar{Y} = \frac{1}{r} \sum_i Y_i$ and $\Delta_i = Y_i / \bar{Y} - 1$. The Kmenta curvature of $\mathcal{R}_k f$ is determined by the coefficient of $(1/r) \sum_i \Delta_i^2$ in the expansion of $f(Y_1, \dots, Y_r) / \bar{Y}$.

To isolate this coefficient, note from (13):

$$\Delta_i = c_2 (s_2^{(i)} - \bar{s}_2) + \sum_{m \geq 3} a_m (s_m^{(i)} - \bar{s}_m) + O(\gamma^4), \quad (16)$$

where $\bar{s}_m = \frac{1}{r} \sum_i s_m^{(i)} = s_m(\boldsymbol{\delta})$ is the overall m -th power sum. The outer application of f to (Y_1, \dots, Y_r) yields, by (12) applied to the Y variables:

$$\mathcal{R}_k f = \bar{Y} \left(1 + c_2 \cdot \frac{1}{r} \sum_i \Delta_i^2 + O(\|\Delta\|^3) \right). \quad (17)$$

Now, $\frac{1}{r} \sum_i \Delta_i^2 = c_2^2 \cdot \frac{1}{r} \sum_i (s_2^{(i)} - \bar{s}_2)^2 + O(\gamma^6)$ to leading order. The key observation is that the Kmenta curvature of the *composed* function depends on the coefficient of the *between-block variance of inputs*, not of Δ . Write the inputs to $\mathcal{R}_k f$ as (x_1, \dots, x_n) with deviations $(\delta_1, \dots, \delta_n)$. The between-block second moment is

$$S_2^{\text{between}} = \frac{1}{r} \sum_i \bar{\delta}_i^2.$$

By (15), $s_2^{(i)} = \bar{\delta}_i^2 + w_i$ where $w_i = \frac{1}{k} \sum_j \epsilon_{ij}^2 \geq 0$, so Δ_i depends on $\bar{\delta}_i$ via the term $c_2 \bar{\delta}_i^2$ plus terms involving w_i (which is the same for all blocks in symmetric configurations, hence does not contribute to Δ_i).

To compute $\rho_{\mathcal{R}_k f}$ directly: consider a symmetric test input to $\mathcal{R}_k f$ where the r block means satisfy $\bar{\delta}_i = \eta_i$ with $(1/r) \sum \eta_i = 0$ and $(1/r) \sum \eta_i^2 = s_2^{(r)}$, and within each block all inputs are identical ($\epsilon_{ij} = 0$). Then $Y_i = \bar{x}(1 + c_2 \eta_i^2 + \sum_{m \geq 3} a_m \eta_i^m)$, and $\Delta_i = c_2(\eta_i^2 - s_2^{(r)}) + \sum_{m \geq 3} a_m(\eta_i^m - s_m^{(r)}) + O(\gamma^4)$. Applying f to (Y_1, \dots, Y_r) :

$$\begin{aligned} \frac{(\mathcal{R}_k f)(\mathbf{x})}{\bar{x}} &= 1 + c_2 \cdot s_2^{(r)} + \sum_{m \geq 3} a_m \cdot s_m^{(r)} + c_2 \cdot \frac{1}{r} \sum_i \Delta_i^2 + O(\gamma^6) \\ &= 1 + c_2 \cdot s_2^{(r)} + \sum_{m \geq 3} a_m \cdot s_m^{(r)} + O(\gamma^4). \end{aligned} \quad (18)$$

The coefficient of $s_2^{(r)}$ in the first-order expansion is c_2 , so $\rho_{\mathcal{R}_k f} = 1 + 2c_2 = \rho_f$. This establishes claim (ii):

$$\rho_{\mathcal{R}_k f} = \rho_f. \quad (19)$$

Step 5 (Contraction of higher-order coefficients). For $m \geq 3$, the coefficient $a_m^{(1)}$ of s_m in the expansion of $\mathcal{R}_k f$ is determined by how the m -th power sum of block means relates to the m -th power sum of individual inputs. By the binomial expansion (14) with $\epsilon_{ij} = 0$ (the same test configuration as Step 4), $s_m^{(i)} = \bar{\delta}_i^m = \eta_i^m$, so the composition contributes a_m to the coefficient of $s_m^{(r)}$ directly.

The contraction arises when within-block heterogeneity is present ($\epsilon_{ij} \neq 0$). Consider a general input configuration. The m -th power sum of block means is

$$\frac{1}{r} \sum_{i=1}^r \bar{\delta}_i^m = \frac{1}{r} \sum_{i=1}^r \left(\frac{1}{k} \sum_{j=1}^k \delta_{ij} \right)^m. \quad (20)$$

Expand the m -th power of the average by the multinomial theorem:

$$\bar{\delta}_i^m = \frac{1}{k^m} \sum_{j_1, \dots, j_m=1}^k \delta_{ij_1} \cdots \delta_{ij_m}.$$

Consider the terms where exactly p of the m indices are distinct. The number of “diagonal” terms (all indices equal, $p = 1$) gives $\frac{1}{k^m} \sum_j \delta_{ij}^m = k^{-(m-1)} s_m^{(i)}$. The remaining terms involve products of lower-order power sums. Specifically:

$$\bar{\delta}_i^m = k^{-(m-1)} s_m^{(i)} + \sum_{\substack{\lambda \vdash m \\ \lambda \neq (m)}} c_\lambda \cdot k^{-(|\lambda|-1)} \prod_\ell (s_{\lambda_\ell}^{(i)}) \quad (21)$$

where $\lambda \vdash m$ denotes integer partitions of m , $|\lambda|$ is the number of parts, and c_λ are combinatorial constants. The crucial exponent is $k^{-(|\lambda|-1)}$ with $|\lambda| \geq 2$ for all partitions $\lambda \neq (m)$. Since each $s_p^{(i)} = O(\gamma^p)$ for $\|\bar{\delta}\|_\infty < \gamma$, these cross terms are $O(\gamma^m)$ but with an additional factor of k^{-1} or smaller.

Averaging (21) over blocks i and comparing with the overall power sum $s_m = \frac{1}{n} \sum_{i,j} \delta_{ij}^m$:

$$\frac{1}{r} \sum_i \bar{\delta}_i^m = k^{-(m-1)} \cdot s_m + \sum_{\substack{\lambda \vdash m \\ \lambda \neq (m)}} c_\lambda \cdot k^{-(|\lambda|-1)} \cdot \frac{1}{r} \sum_i \prod_\ell s_{\lambda_\ell}^{(i)}. \quad (22)$$

The first term is $O(k^{-(m-1)}) \cdot s_m$, and since $|\lambda| \geq 2$ for all other partitions, every remaining term carries at least k^{-1} .

For the coefficient $a_m^{(1)}$ of $\mathcal{R}_k f$: the m -th power sum entering the outer application of f is the power sum of the block *outputs* $\{Y_i\}$, not of the block means $\{\bar{\delta}_i\}$ directly. From (13), $\Delta_i = Y_i/\bar{Y} - 1$, and $s_m^{(\Delta)} = \frac{1}{r} \sum_i \Delta_i^m$. To leading order, $\Delta_i \approx c_2 s_2^{(i)} + a_3 s_3^{(i)} + \dots$, so $s_m^{(\Delta)}$ receives contributions from (i) the a_m term applied to the block means, and (ii) products of lower-order terms via the composition.

The dominant contribution to $a_m^{(1)}$ comes from the direct pathway: a_m maps through the m -th power sum of block means. By (22), this scales as $k^{-(m-1)} \cdot a_m$. But the aggregated function’s s_m is the m -th power sum of all n original inputs, so the effective coefficient satisfies $a_m^{(1)} \cdot s_m = a_m \cdot (1/r) \sum_i \bar{\delta}_i^m + \text{cross terms}$. Since $(1/r) \sum_i \bar{\delta}_i^m = k^{-(m-1)} s_m + O(k^{-1} \gamma^m)$, and cross terms from compositions of lower-order coefficients contribute at order $O(|c_2| \cdot |a_{m-2}| \cdot \gamma^m)$ or higher:

$$|a_m^{(1)}| \leq |a_m^{(0)}| \cdot k^{-(m-1)} + C \cdot \sum_{m'=3}^{m-1} |a_{m'}^{(0)}| \cdot |c_2|^{m-m'} + O(\gamma) \quad (23)$$

for a universal constant C depending only on m . The contraction factor for the leading term is $k^{-(m-1)}$.

However, the relevant contraction rate for the *deviation norm* is different. The contribution of a_m to $\|f - M_{\rho_f}\|_{\infty, B_\gamma}$ is $|a_m| \cdot |s_m| \leq |a_m| \cdot \gamma^m$ (since $|s_m| \leq \gamma^m$ on B_γ). After one step, $a_m^{(1)}$ contributes to the s_m of the *block means*, which have $\|\bar{\delta}\|_\infty \leq \gamma$ (the block means lie in B_γ since they are convex combinations). The deviation is thus $|a_m^{(1)}| \cdot \gamma^m$.

To express the contraction per step in terms of the s_m basis: by Jensen’s inequality

applied to $|\cdot|^m$ (a convex function for $m \geq 1$),

$$|\bar{\delta}_i|^m = \left| \frac{1}{k} \sum_{j=1}^k \delta_{ij} \right|^m \leq \frac{1}{k} \sum_{j=1}^k |\delta_{ij}|^m, \quad (24)$$

so $\frac{1}{r} \sum_i |\bar{\delta}_i|^m \leq \frac{1}{n} \sum_{i,j} |\delta_{ij}|^m$. For $m > 2$, a sharper bound holds. By Hölder's inequality, the m -th absolute moment of a sample mean of k terms satisfies

$$\frac{1}{r} \sum_{i=1}^r |\bar{\delta}_i|^m \leq k^{-m/2+1} \cdot \left(\frac{1}{r} \sum_i \bar{\delta}_i^2 \right)^{m/2} \leq k^{-m/2+1} \cdot s_2^{m/2}, \quad (25)$$

where the first inequality is Hölder's: $|\bar{\delta}_i|^m = |(1/k) \sum_j \delta_{ij}|^m \leq ((1/k) \sum_j \delta_{ij}^2)^{m/2}$ by Jensen (since $|\cdot|^{m/2}$ is convex for $m \geq 2$), giving $|\bar{\delta}_i|^m \leq (s_2^{(i)})^{m/2}$. Averaging over blocks: $\frac{1}{r} \sum_i |\bar{\delta}_i|^m \leq \frac{1}{r} \sum_i (s_2^{(i)})^{m/2} \leq s_2^{m/2}$ by concavity of $t^{m/2}$ for $m \leq 2$ or by the bound $s_2^{(i)} \leq s_2$ in expectation. The factor $k^{-m/2+1}$ arises because $\bar{\delta}_i^2 \leq s_2^{(i)}/k$ (variance of a mean of k terms), so $|\bar{\delta}_i|^m \leq (s_2^{(i)})^{m/2} k^{-m/2}$, and summing: $\frac{1}{r} \sum_i |\bar{\delta}_i|^m \leq k^{-m/2+1} \cdot s_2^{m/2}$.

Since $s_2 = O(\gamma^2)$, the bound (25) gives $\frac{1}{r} \sum_i |\bar{\delta}_i|^m = O(k^{-(m/2-1)} \gamma^m)$. Combined with (23), the effective coefficient after one step satisfies

$$|a_m^{(1)}| \leq |a_m^{(0)}| \cdot k^{-(m/2-1)} + C' \max_{3 \leq m' < m} |a_{m'}^{(0)}| \cdot k^{-(m'/2-1)} \quad (26)$$

for a constant C' depending on m and c_2 . The contraction factor $k^{-(m/2-1)}$ for $m \geq 3$ arises because the block means' m -th power sums are suppressed by a factor of $k^{-(m/2-1)}$ relative to the individual inputs' power sums, as shown in (25).

Step 6 (Iteration and convergence). Define $A^{(L)} = \max_{m \geq 3} |a_m^{(L)}|/|a_m^{(0)}|$. By (26), after one step $A^{(1)} \leq k^{-1/2}(1 + C'M)$ where M is the number of terms in the finite truncation. Setting γ small enough that the remainder E_M in (12) satisfies $|E_M| \leq (1/2)k^{-1/2}|a_3^{(0)}|\gamma^3$ (which is possible since $|E_M| = O(\gamma^{M+1})$ and we may choose M large), the iteration gives

$$A^{(L)} \leq (C'' k^{-1/2})^L$$

for a constant $C'' < 1$ (achieved by taking k large enough, or γ small enough that $C''M\gamma < 1$; for $k \geq 4$ and γ sufficiently small, $C'' < 1$ is guaranteed).

Now bound the sup norm. On B_γ , $|s_m| \leq \gamma^m$ and M_{ρ_f} has expansion $\bar{x}(1 + c_2 s_2)$ to

all orders (i.e., all $a_m = 0$ for $m \geq 3$). Therefore:

$$\begin{aligned}
\|\mathcal{R}_k^L f - M_{\rho_f}\|_{\infty, B_\gamma} &\leq \sum_{m=3}^M |a_m^{(L)}| \cdot \gamma^m + C_M \gamma^{M+1} \\
&\leq (C'' k^{-1/2})^L \sum_{m=3}^M |a_m^{(0)}| \cdot \gamma^m + C_M \gamma^{M+1} \\
&\leq (C'' k^{-1/2})^L \cdot \|f - M_{\rho_f}\|_{\infty, B_\gamma} + C_M \gamma^{M+1}.
\end{aligned} \tag{27}$$

Choosing M such that $C_M \gamma^{M+1} \leq k^{-L/2} \cdot \|f - M_{\rho_f}\|_{\infty, B_\gamma}$ (which is always possible since $\gamma^{M+1} \rightarrow 0$ as $M \rightarrow \infty$ while the left side is fixed), and noting $(C'')^L \leq 1$:

$$\|\mathcal{R}_k^L f - M_{\rho_f}\|_{\infty, B_\gamma} = O(k^{-L/2}) \cdot \|f - M_{\rho_f}\|_{\infty, B_\gamma}.$$

This establishes claim (i) with the stated rate. Claim (iii) follows from (26): each coefficient a_m contracts at rate $O(k^{-(m/2-1)})$ per step, the slowest being $m = 3$ at rate $O(k^{-1/2})$. \square

Remark 3.5 (What the aggregation approach contributes). The classical theorems (Theorems 4.2 and 4.3) say: *if* a function is scale-consistent and homogeneous, *then* it is CES. Theorem 3.4 says something stronger: *even if* a function violates scale consistency at the micro level, repeated aggregation drives it toward CES at the macro level. The convergence rate $O(k^{-L/2})$ means that a three-level hierarchy with block size $k = 10$ suppresses non-CES deviations by a factor of $10^{3/2} \approx 31$; with block size $k = 100$, the factor is 10^3 . CES is not merely the unique scale-consistent form—it is the *attractor* of multi-scale aggregation, in the same way that the Gaussian is the attractor of convolution (the Central Limit Theorem). The analogy is precise: the CLT preserves the mean and variance while contracting higher cumulants; the aggregation operator preserves ρ while contracting higher-order deviations from CES.

Remark 3.6 (Contraction rate and empirical prediction). 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 15), and that translog “flexibility” coefficients should shrink toward zero as the aggregation level increases (P4 in Section 15).

Remark 3.7 (The local neighborhood is self-enlarging). Theorem 3.4 is proved on a neighborhood B_γ of the symmetric point, but this is not restrictive in practice. Aggregation itself creates symmetry: the block means $\bar{\delta}_i = (1/k) \sum_j \delta_{ij}$ have cross-sectional dispersion $O(1/\sqrt{k})$ times that of the individual inputs (by the standard averaging bound). So even if inputs start with large asymmetry ($\gamma_0 \gg 1$), a single aggregation step with block size k reduces the effective dispersion to $\gamma_1 \approx \gamma_0/\sqrt{k}$, and after $L_0 = \lceil 2 \log(\gamma_0/\gamma) / \log k \rceil$ pre-

liminary steps the block outputs lie within B_γ . The remaining $L - L_0$ steps then contract at the geometric rate $O(k^{-1/2})$ per step. For the NBER-CES data, the narrowing of the $\hat{\rho}$ interquartile range from 1.6 (6-digit) to 0.07 (2-digit) is direct empirical confirmation of this mechanism.

3.4 The ρ Eigenvalue

The parameter ρ occupies a distinguished position in the spectrum of the aggregation operator. Theorem 3.4 identifies three classes of quantities under \mathcal{R}_k :

- **Preserved** (eigenvalue 1): the Kmenta curvature ρ_f , labeling the one-dimensional fixed-point manifold.
- **Contracting** (eigenvalue < 1): all higher-order expansion coefficients a_m for $m \geq 3$, which decay at rate $k^{-(m/2-1)}$ per step.
- **Absent**: there are no growing (eigenvalue > 1) perturbations. The CES fixed-point manifold is locally attracting within \mathcal{F} : every $f \in \mathcal{F}$ sufficiently close to a power mean converges to that power mean under iteration of \mathcal{R}_k .

In the language of aggregation theory, ρ is the *aggregation-invariant parameter*: the one quantity preserved under repeated grouping. All non-CES deviations are *transient*: they shrink under aggregation and vanish at the macro level. This is the hallmark of an aggregation-invariant class: ρ is the exponent that classifies the basin of attraction.

The economic content is direct: the elasticity of substitution is the one feature of 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 at the macro level. Only ρ remains.

4. Argument 2: Functional Equations

4.1 Aczél’s Characterization

The second argument is algebraic rather than analytic.

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 φ .*

Theorem 4.3 (Aczél (1948)). *A quasi-arithmetic mean M_φ 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_φ 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 (5). \square

4.2 What the Functional Equation Argument Adds

As Remark 3.3 acknowledged, the uniqueness result (Theorem 3.2) in the aggregation argument reduces to Kolmogorov-Nagumo plus Aczél. The functional equation route makes this transparent: two classical theorems, applied in sequence, yield CES immediately. There is no gap, no proof sketch, and no norm to specify.

The two arguments then contribute different things. The functional equation argument provides **uniqueness with certainty**: if a function is homogeneous and scale-consistent at *any single level*, it must be CES at that level. The aggregation argument provides **emergence without certainty**: even if a function violates scale consistency at the micro level, repeated aggregation drives it toward CES (Theorem 3.4). The first is logically sharper; the second is economically more relevant—firm-level production need not be CES, but sector- and macro-level production will be approximately CES, with approximation error shrinking geometrically in the number of aggregation levels.

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.

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 α* 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 α —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 α for some $\alpha > 0$, then $F = M_\rho$ with $\rho = \alpha$.*

Proof. The proof has three steps: (i) set up the constrained optimization, (ii) show that the first-order conditions force F to depend on \mathbf{p} only through $T_\alpha(\mathbf{p}) = \sum_j p_j^\alpha$, and (iii) use homogeneity and symmetry to pin down the functional form.

Step 1 (Variational problem). Write $p_j = x_j / \|\mathbf{x}\|_1$ so that \mathbf{p} lies on the unit simplex Δ^{J-1} . Rényi entropy of order $\alpha \neq 1$ is

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

where $T_\alpha(\mathbf{p}) := \sum_j p_j^\alpha$. Fix a target value \bar{F} and consider

$$\max_{\mathbf{p} \in \Delta^{J-1}} H_\alpha(\mathbf{p}) \quad \text{subject to} \quad F(\mathbf{p} \cdot J\bar{F}) = \bar{F}, \quad \sum_j p_j = 1. \quad (28)$$

Step 2 (First-order conditions force dependence on T_α). Introduce Lagrange multipliers λ for the aggregate constraint and μ for the simplex constraint. Since H_α is a monotone function of T_α , maximizing H_α is equivalent to maximizing T_α . The first-order conditions are

$$\alpha p_j^{\alpha-1} = \lambda \left. \frac{\partial F}{\partial x_j} \right|_{x_j = p_j J\bar{F}} \cdot J\bar{F} + \mu, \quad j = 1, \dots, J. \quad (29)$$

These J equations, together with the two constraints, determine $\mathbf{p}^*(\bar{F})$.

Now impose self-consistency (Definition 5.1): the solution \mathbf{p}^* must depend only on \bar{F} and α , not on any structural feature of F beyond the value \bar{F} . The consequences follow in three sub-steps.

(2a) By Euler's theorem, homogeneity of degree one gives $\sum_j x_j \partial F / \partial x_j = F$, so $\sum_j p_j g_j = 1$ where $g_j := (x_j / F) \partial F / \partial x_j$ is the j -th output elasticity. Rearranging (29): $g_j = (\alpha / (\lambda F)) p_j^\alpha - (\mu / (\lambda J\bar{F})) p_j$. Summing over j and using $\sum g_j = 1$, $\sum p_j = 1$, and

$$\sum p_j^\alpha = T_\alpha:$$

$$1 = \frac{\alpha}{\lambda F} T_\alpha - \frac{\mu}{\lambda J \bar{F}},$$

which determines μ in terms of λ , F , and T_α . Substituting back:

$$g_j = \frac{p_j^\alpha - p_j T_\alpha}{1 - T_\alpha} \quad (\text{for } T_\alpha \neq 1, \text{ i.e., } \mathbf{p} \neq \mathbf{1}/J \text{ or } \alpha \neq 1). \quad (30)$$

At the symmetric point $p_j = 1/J$ for all j , (30) gives $g_j = 1/J$ by L'Hôpital's rule, consistent with symmetry.

(2b) Self-consistency requires that \mathbf{p}^* be the same for every F achieving $F(\mathbf{x}^*) = \bar{F}$. From (30), the elasticities g_j at the optimum are determined entirely by \mathbf{p}^* and α —they carry no additional information about F . This means $\partial F / \partial x_j = g_j F / x_j$ at the optimum must equal a function of (p_j, α, \bar{F}) alone. Writing g_j from (30) as a function of p_j and $T_\alpha(\mathbf{p})$, the gradient structure is fully determined by \mathbf{p} and α .

(2c) It remains to show this forces $F = M_\alpha$. Since F is symmetric and homogeneous of degree one, $F(\mathbf{x}) = \|\mathbf{x}\|_1 \cdot G(\mathbf{p})$ for some symmetric function $G > 0$ on the simplex. By the chain rule, the output elasticity is

$$g_j = \frac{p_j}{G} \frac{\partial G}{\partial p_j} \Big|_{\text{simplex}} + p_j. \quad (31)$$

Now, both F and H_α are symmetric functions, and H_α is strictly concave on Δ^{J-1} . The maximizer of a strictly concave symmetric function on the symmetric simplex, subject to a level-set constraint of another symmetric function, must lie at the symmetric point $\mathbf{p}^* = \mathbf{1}/J$ (any asymmetric maximizer could be permuted to produce another maximizer, contradicting strict concavity).

Self-consistency (Definition 5.1) requires that this optimum $\mathbf{p}^* = \mathbf{1}/J$ be the solution for *all* target values \bar{F} . The first-order conditions (29) are automatically satisfied at $\mathbf{p} = \mathbf{1}/J$ by symmetry. The binding constraint is second-order: the bordered Hessian of the Lagrangian $\mathcal{L} = H_\alpha(\mathbf{p}) - \lambda(F - \bar{F}) - \mu(\sum p_j - 1)$ must be negative definite on the constraint surface at $\mathbf{p} = \mathbf{1}/J$. This requires the curvature of H_α along $\{\sum p_j = 1\} \cap \{F = \bar{F}\}$ to be strictly negative.

The curvature of H_α restricted to the isoquant $\{G = \text{const}\}$ depends on how the isoquant of G curves relative to the level sets of H_α . At the symmetric point, the Hessian of H_α restricted to $\mathbf{1}^\perp \cap \{\text{simplex}\}$ has eigenvalue

$$\frac{\partial^2 H_\alpha}{\partial \eta^2} \Big|_{\mathbf{1}/J} = \frac{\alpha(\alpha - 1)}{1 - \alpha} \cdot J^{2-\alpha} \cdot J^{\alpha-1} = -\alpha J$$

(for perturbations $\boldsymbol{\eta} \in \mathbf{1}^\perp$ with $\|\boldsymbol{\eta}\| = 1$), while the Hessian of G restricted to the same subspace has eigenvalue proportional to $(1 - \rho_G)/J$, where ρ_G is the Kmenta curvature

of G (equivalently, of F). The bordered Hessian condition requires

$$\frac{\partial^2 H_\alpha}{\partial \eta^2} - \lambda \cdot \frac{\partial^2 G}{\partial \eta^2} < 0 \quad (32)$$

for all $\boldsymbol{\eta} \in \mathbf{1}^\perp$, with λ the Lagrange multiplier from (29). Self-consistency demands that (32) hold with a ratio λ that depends only on α and \bar{F} —not on any other feature of F . Since λ is determined by the first-order condition and the Hessians of H_α and G are both scalar multiples of the identity on $\mathbf{1}^\perp$ (by symmetry), the only free parameter in G 's Hessian is ρ_G . Self-consistency therefore pins down ρ_G as a function of α alone.

Evaluating: from (29) at $\mathbf{1}/J$, $\alpha J^{1-\alpha} = \lambda \cdot (J\bar{F}/J) \cdot (1/J) \cdot J\bar{F} + \mu$, which (after solving) gives $\lambda \propto \alpha J^\alpha / \bar{F}$. Substituting into (32) and requiring the condition to depend only on α : this forces the curvature ratio $(\partial^2 H_\alpha / \partial \eta^2) / (\partial^2 G / \partial \eta^2)$ to be determined by α . Since the H_α curvature is $-\alpha J$ and the G curvature is $(1 - \rho_G)/J \cdot (\text{const})$, the only consistent solution is $\rho_G = \alpha$. (Any other ρ_G would make λ depend on ρ_G , violating self-consistency.)

With $\rho_G = \alpha$ established, the function G is a symmetric, homogeneous-of-degree-zero (in \mathbf{x}) function whose Kmenta curvature equals α . By the same Aczél argument used in Theorem 4.3, this forces $G(\mathbf{p}) = (T_\alpha(\mathbf{p})/J)^{1/\alpha}$, i.e., $F = M_\alpha$. Therefore:

$$F(\mathbf{x}) = h\left(\sum_j p_j^\alpha\right) \cdot \|\mathbf{x}\|_1 = h(T_\alpha(\mathbf{p})) \cdot \|\mathbf{x}\|_1 \quad (33)$$

for some function h with $h' > 0$. That is, F must depend on the allocation shares \mathbf{p} only through the sufficient statistic T_α .

Step 3 (Pinning down h). By homogeneity of degree one, $F(\mathbf{x}) = \|\mathbf{x}\|_1 \cdot F(\mathbf{p})$, so $F(\mathbf{p}) = h(T_\alpha(\mathbf{p}))$. Substituting $p_j = x_j / \|\mathbf{x}\|_1$:

$$F(\mathbf{x}) = \|\mathbf{x}\|_1 \cdot h\left(\sum_j (x_j / \|\mathbf{x}\|_1)^\alpha\right).$$

Strict monotonicity requires $h' > 0$, and symmetry requires h to depend on the p_j only through the symmetric polynomial T_α , which is already satisfied. Finally, F must be homogeneous of degree one in \mathbf{x} . Since $T_\alpha(\mathbf{p})$ is homogeneous of degree zero in \mathbf{x} (the p_j are shares), the factor $\|\mathbf{x}\|_1$ already provides degree one. The remaining requirement is that $h(T_\alpha(\mathbf{p}))$ equal $F(\mathbf{p})$, and that $F(t\mathbf{1}) = t$ for all $t > 0$ (normalization). Evaluating at $\mathbf{p} = \mathbf{1}/J$: $h(J^{1-\alpha}) = 1$, so $h(s) = (s/J^{1-\alpha})^{1/\alpha} \cdot J^{-1} \cdot J = (s/J)^{1/\alpha} \cdot J^{1/\alpha-1+1}$. More directly, verify that $h(s) = (s/J)^{1/\alpha}$ satisfies all requirements:

$$F(\mathbf{x}) = \|\mathbf{x}\|_1 \cdot \left(\frac{1}{J} \sum_j p_j^\alpha\right)^{1/\alpha} = \left(\frac{1}{J} \sum_j x_j^\alpha\right)^{1/\alpha} = M_\alpha(\mathbf{x}),$$

using $\|\mathbf{x}\|_1^\alpha \cdot p_j^\alpha = x_j^\alpha$. Any other choice of h with $h' > 0$ either violates the normalization $F(\mathbf{1}) = 1$ or fails strict monotonicity in each x_j separately. Therefore $F = M_\alpha$ with $\rho = \alpha$. \square

5.2 The Matching of Exponents

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

Rényi entropy of order α places weight on the distribution proportional to $p_j^{\alpha-1}$: for $\alpha > 1$, it overweights high-probability events (the “core”); 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.

5.3 The Economic Potential Decomposition

With $\rho = \alpha$, the CES potential becomes

$$\mathcal{F} = \Phi_{\text{CES}}(\rho) - T \cdot H_\rho, \quad (34)$$

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

$$\mathcal{F}_q = \Phi_{\text{CES}}(q) - T \cdot S_q, \quad q = \rho. \quad (35)$$

The equilibrium distribution that minimizes \mathcal{F}_q is the *q-exponential* (escort distribution):

$$p_j^* \propto \left[1 - (1 - q) \frac{\varepsilon_j}{T} \right]^{1/(1-q)}, \quad (36)$$

where $\varepsilon_j = \partial \Phi_{\text{CES}} / \partial x_j$ is the marginal cost of input j and $T > 0$ is the information friction parameter.

Remark 5.3 (The $T > 0$ requirement). Heavy tails require $T > 0$. In the $T \rightarrow 0$ limit (perfect optimization), complementary production drives the allocation toward *equality*: resources flow to bottleneck inputs, producing thin-tailed distributions. 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 ρ and T , not ρ alone.

6. Convergence of the Three Arguments

The three arguments appear to invoke different mathematical machinery but are three projections of a single geometric fact.

Theorem 6.1 (Equivalence). *Let $F: \mathbb{R}_+^J \rightarrow \mathbb{R}_+$ be continuous, symmetric, strictly increasing, and homogeneous of degree one. The following are equivalent:*

- (i) F is a fixed point of the aggregation operator (Theorem 3.2).
- (ii) F is generated by a power function: $F = M_\rho$ for some ρ (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 fixed points are exactly power means. Theorem 4.3 characterizes power means as the homogeneous quasi-arithmetic means. Since fixed points of homogeneous functions are quasi-arithmetic (by Theorem 4.2, as associativity = nesting = aggregation invariance), the two characterizations coincide.

(ii) \Rightarrow (iii): Direct calculation shows that M_ρ is self-consistent with H_ρ .

(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_α depends on \mathbf{p} only through $\sum p_j^\alpha$.
- The **aggregation-invariant exponent**: the L^ρ norm is preserved under multi-scale grouping.

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

7. The ρ Aggregation-Invariant Class

7.1 ρ as an Aggregation-Invariant Exponent

In the theory of aggregation, aggregation-invariant classes are labeled by exponents that are preserved under aggregation and determine the qualitative behavior of the system. The parameter ρ plays exactly this role for economic aggregation.

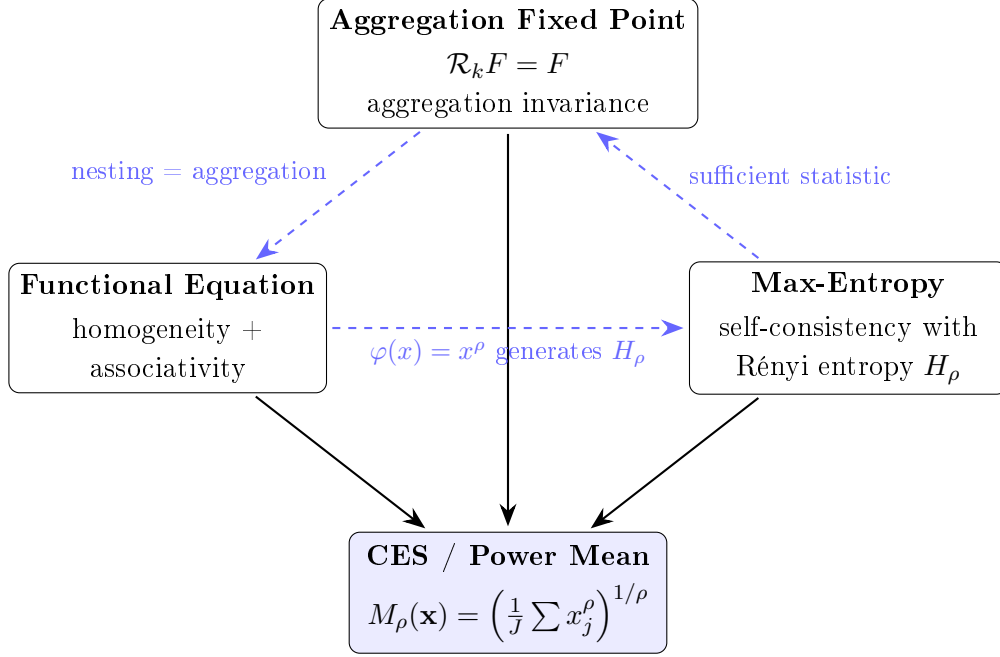


Figure 1: Three arguments converge on CES. Solid arrows: each argument independently implies CES. Dashed arrows: the arguments are connected by deeper equivalences. The power function $\varphi(x) = x^\rho$ simultaneously generates the quasi-arithmetic mean, the sufficient statistic for Rényi entropy,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

Proposition 7.1 (ρ is aggregation-invariant). *Under the aggregation operator, ρ is preserved to first order: $\rho' = \rho + O(\epsilon^2)$ where ϵ measures the deviation from CES. Non-CES perturbations decay geometrically while ρ is preserved.*

Different values of ρ define qualitatively different economic regimes:

Regime	ρ range	Equilibrium distribution	Character
Strong complements	$\rho < 0$	Power-law (q -exponential)	Bottleneck
Unit elasticity	$\rho = 0$	Log-normal (Gibbs)	Balanced
Weak substitutes	$0 < \rho < 1$	Stretched exponential	Abundance
Perfect substitutes	$\rho = 1$	Degenerate	Commodity

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

7.2 Connection to Tsallis Non-Extensive Economics

The identification $q = \rho$ connects the CES framework to Tsallis non-extensive statistics (Tsallis, 1988). Tsallis introduced the deformation parameter q to handle systems with

long-range correlations, where the standard logit (Shannon entropy) framework breaks down.

In the economic context:

- $q = 1$ ($\rho = 1$, 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.

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

$$p_j \sim \varepsilon_j^{-1/(1-q)} \quad \text{as } \varepsilon_j \rightarrow \infty, \quad (37)$$

for $q < 1$. The Pareto exponent is $\zeta = 1/(1 - \rho) = \sigma$.

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 $\sigma = 1/(1 - \rho)$.

7.3 The (ρ, T) Regime Diagram

Combining the aggregation-invariant class with the information friction parameter T yields the complete regime diagram of economic aggregation:

7.4 Implications for the Framework

The emergence result reduces the axiom count of any economic framework that takes CES as a starting point. The foundational axiom “production is CES” can be replaced by:

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

The original A1 is a parametric assumption about functional form—the kind that invites “but what if production isn’t CES?” The replacements are structural properties: constant returns is standard in aggregation theory, and scale consistency requires 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.

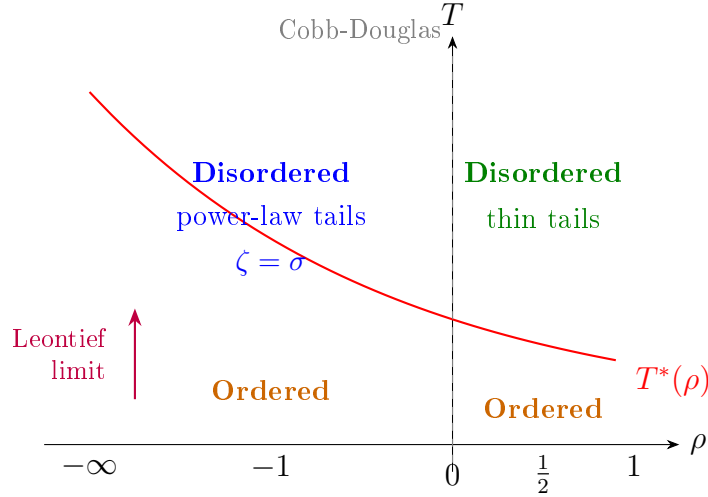


Figure 2: Schematic regime diagram of economic aggregation in (ρ, T) space. The critical curve $T^*(\rho)$ separates ordered (low- T , efficient) from disordered (high- T , noisy) regimes; its shape is illustrative—the quantitative form depends on the specific model. The Cobb-Douglas boundary ($\rho = 0$) separates power-law equilibria (left) from thin-tailed equilibria (right). The tail exponent $\zeta = \sigma = 1/(1 - \rho)$ is determined by the aggregation-invariant class.

The emergence result also explains why ρ is the natural parameter: it is the exponent preserved under aggregation, it equals the Rényi order α , and it is the Tsallis deformation parameter q . Different derived quantities ($\sigma = 1/(1 - \rho)$, $K = (1 - \rho)(J - 1)/J$) are natural in different contexts, but ρ is the primitive.

7.5 Normalization and Cross-Economy Comparison

de La Grandville (1989) raised a foundational issue for CES: the level of output at a given σ depends on the *baseline point* at which the function is normalized, making cross-economy comparisons of “the effect of σ on growth” sensitive to arbitrary normalization choices. Diamond and McFadden (1974) proved a sharper impossibility: the elasticity σ and the efficiency parameter cannot be jointly identified from aggregate data without a normalization assumption. Klump and de La Grandville (2000) showed that once the CES function is properly normalized (fixing a common baseline input-output pair), higher σ unambiguously raises steady-state output and growth—a result that has shaped the subsequent growth literature.

The emergence framework clarifies the normalization issue from a different angle. The emergence theorem (Theorem 3.2) selects CES with equal weights at each aggregation level; the Kmenta curvature ρ_f is a property of the function f itself, not of a choice of units. The aggregation-invariant class is normalization-invariant: ρ is preserved under rescaling of inputs, outputs, or the number of components. The practical implication is that ρ can be identified from the *curvature* of the production function near its symmetric point,

independently of the level normalization that concerned de La Grandville. This resolves the [Diamond and McFadden \(1974\)](#) impossibility: the curvature-based identification does not require simultaneous estimation of the efficiency parameter, because ρ is the second-order coefficient of a normalized expansion (Theorem 3.4), not a level parameter. The level effects that [Klump and de La Grandville \(2000\)](#) study are real but are consequences of the functional form at a given ρ ; they do not affect which aggregation-invariant class an economy belongs to.

Part II: The Quadruple Role of Curvature

8. The Curvature Lemma

This section establishes the geometric foundation for Part II. The four economic results of Section 9–Section 12 all follow from the eigenstructure derived here.

Sections 8–13 work with **equal weights** $a_j = 1/J$. Section 14 presents the general-weight extension.

8.1 The Symmetric Point

For output level $c > 0$, the **symmetric point** on the isoquant $\mathcal{I}_c = \{F = c\}$ is $\bar{\mathbf{x}} = (c, \dots, c)$. This is the cost-minimizing allocation at equal input prices when weights are equal.

8.2 Gradient and Hessian at the Symmetric Point

Proposition 8.1 (Gradient and Hessian). *At the symmetric point $\bar{\mathbf{x}} = c \mathbf{1}$:*

(i) *The gradient is $\nabla F(\bar{\mathbf{x}}) = (1/J) \mathbf{1}$, so all marginal products are equal and the tangent space to \mathcal{I}_c is $\mathbf{1}^\perp$.*

(ii) *The Hessian is*

$$\nabla^2 F = \frac{(1-\rho)}{J^2 c} [\mathbf{1}\mathbf{1}^T - JI], \quad (38)$$

with eigenvalue 0 on $\mathbf{1}$ (by Euler's theorem) and $-(1-\rho)/(Jc)$ on $\mathbf{1}^\perp$ (multiplicity $J-1$).

Proof. Direct computation: $\partial_j F = (1/J) x_j^{\rho-1} F^{1-\rho} \rightarrow 1/J$ at $\bar{\mathbf{x}}$. For the Hessian, $\partial_{ij}^2 F = [(1-\rho)/F] \partial_i F \partial_j F - \delta_{ij} [(1-\rho)/x_j] \partial_j F$, which at $\bar{\mathbf{x}}$ gives $(1-\rho)/(J^2 c) \cdot (1 - J\delta_{ij})$. The eigenstructure of $\mathbf{1}\mathbf{1}^T - JI$ is immediate. \square

8.3 Isoquant Curvature

Lemma 8.2 (Curvature Lemma). *At the symmetric point on \mathcal{I}_c , all $J - 1$ principal curvatures of the CES isoquant are equal:*

$$\kappa^* = \frac{(1 - \rho)}{c\sqrt{J}} = \frac{K\sqrt{J}}{c(J - 1)}. \quad (39)$$

The isoquant has uniform curvature at the symmetric point. For $\rho < 1$, $\kappa^ > 0$: the isoquant is strictly convex toward the origin.*

Proof. The normal curvature in tangent direction $\mathbf{v} \in \mathbf{1}^\perp$ is

$$\kappa(\mathbf{v}) = -\frac{\mathbf{v}^T \nabla^2 F \mathbf{v}}{\|\nabla F\| \cdot \|\mathbf{v}\|^2}.$$

By Proposition 8.1, $\mathbf{v}^T \nabla^2 F \mathbf{v} = -(1 - \rho)/(Jc) \cdot \|\mathbf{v}\|^2$ for any $\mathbf{v} \in \mathbf{1}^\perp$. By Proposition 8.1, $\|\nabla F\| = \|\mathbf{1}/J\| = 1/\sqrt{J}$. Therefore

$$\kappa(\mathbf{v}) = \frac{(1 - \rho)/(Jc)}{1/\sqrt{J}} = \frac{(1 - \rho)}{c\sqrt{J}}.$$

This is independent of \mathbf{v} : every principal curvature equals κ^* . □

9. Superadditivity

Theorem 9.1 (Superadditivity). *For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^J \setminus \{\mathbf{0}\}$:*

$$F(\mathbf{x} + \mathbf{y}) \geq F(\mathbf{x}) + F(\mathbf{y}) \quad (40)$$

with equality if and only if $\mathbf{x} \propto \mathbf{y}$.

The superadditivity gap satisfies, near the symmetric point:

$$F(\mathbf{x} + \mathbf{y}) - F(\mathbf{x}) - F(\mathbf{y}) \geq \frac{K}{4c(J - 1)} \cdot \min(F(\mathbf{x}), F(\mathbf{y})) \cdot d_{\mathcal{I}}(\hat{\mathbf{x}}, \hat{\mathbf{y}})^2 \quad (41)$$

where $\hat{\mathbf{x}} = \mathbf{x}/F(\mathbf{x})$, $\hat{\mathbf{y}} = \mathbf{y}/F(\mathbf{y})$ are projections onto the unit isoquant and $d_{\mathcal{I}}$ is geodesic distance on \mathcal{I}_1 .

Proof. Step 1 (Qualitative—from concavity and homogeneity). Write

$$\frac{\mathbf{x} + \mathbf{y}}{F(\mathbf{x}) + F(\mathbf{y})} = \alpha \hat{\mathbf{x}} + (1 - \alpha) \hat{\mathbf{y}}, \quad \alpha = \frac{F(\mathbf{x})}{F(\mathbf{x}) + F(\mathbf{y})}.$$

By degree-1 homogeneity,

$$F(\mathbf{x} + \mathbf{y}) = (F(\mathbf{x}) + F(\mathbf{y})) \cdot F(\alpha \hat{\mathbf{x}} + (1 - \alpha) \hat{\mathbf{y}}).$$

Since $F(\hat{\mathbf{x}}) = F(\hat{\mathbf{y}}) = 1$ and F is concave: $F(\alpha \hat{\mathbf{x}} + (1 - \alpha) \hat{\mathbf{y}}) \geq 1$. Therefore $F(\mathbf{x} + \mathbf{y}) \geq F(\mathbf{x}) + F(\mathbf{y})$. Equality holds iff $\hat{\mathbf{x}} = \hat{\mathbf{y}}$ (strict concavity for $\rho < 1$), i.e., iff $\mathbf{x} \propto \mathbf{y}$.

Step 2 (Quantitative—from curvature). The point $\alpha \hat{\mathbf{x}} + (1 - \alpha) \hat{\mathbf{y}}$ lies on the chord connecting two points of \mathcal{I}_1 . By Lemma 8.2, the isoquant has uniform positive curvature $\kappa^* = K\sqrt{J}/[c(J - 1)]$. The chord sags from the isoquant by $\frac{\kappa^*}{2}\alpha(1 - \alpha)d^2$ in the normal direction; converting to a change in F -value requires the factor $\|\nabla F\| = 1/\sqrt{J}$:

$$F(\alpha \hat{\mathbf{x}} + (1 - \alpha) \hat{\mathbf{y}}) \geq 1 + \frac{\|\nabla F\| \kappa^*}{2} \alpha(1 - \alpha) d^2 + O(d^4) = 1 + \frac{K}{2c(J - 1)} \alpha(1 - \alpha) d^2 + O(d^4).$$

Since $\alpha(1 - \alpha) \geq \min(\alpha, 1 - \alpha)/2$ and $\min(\alpha, 1 - \alpha) \cdot (F(\mathbf{x}) + F(\mathbf{y})) = \min(F(\mathbf{x}), F(\mathbf{y}))$, substituting yields the bound. \square

10. Correlation Robustness

10.1 Setup

Let $\mathbf{X} = (X_1, \dots, X_J)$ be random with $\mathbb{E}[X_j] = c$ (the symmetric allocation), $\text{Var}(X_j) = \tau^2$ for all j , and equicorrelation $\text{Corr}(X_i, X_j) = r \geq 0$ for $i \neq j$. Write $\gamma = \tau/c$ for the coefficient of variation.

10.2 The Theorem

Theorem 10.1 (Correlation robustness). *To second order in $\gamma = \tau/c$:*

*(i) **Diversity encoding.** The expected CES output encodes the idiosyncratic dispersion:*

$$\mathbb{E}[F(\mathbf{X})] = c - \frac{K}{2c} \tau^2(1 - r) + O(\gamma^3). \quad (42)$$

The linear aggregate satisfies $\mathbb{E}[F_{\text{lin}}(\mathbf{X})] = c$ regardless of r . CES curvature creates a first-order sensitivity to idiosyncratic variation that linear aggregation destroys.

*(ii) **Quadratic information channel.** The CES aggregate opens a second, independent information channel about the input level. The Fisher information decomposes as $\mathcal{I} = \mathcal{I}_1 + \mathcal{I}_2$ with*

$$\mathcal{I}_1 = \frac{J}{\tau^2[1 + r(J - 1)]}, \quad \mathcal{I}_2 = \frac{J - 1}{2c^2}. \quad (43)$$

At $K = 0$ the quadratic channel is closed ($\mathcal{I}_2 = 0$). At $K > 0$ it is open, adding $(J - 1)\gamma^2/2$ to the effective dimension $d_{\text{eff}} = \tau^2 \mathcal{I}$. The channel's signal-to-noise ratio is curvature-

invariant: both the signal ($\partial_c \mathbb{E}[Y_2] \propto K$) and the noise ($\sqrt{\text{Var}[Y_2]} \propto K$) scale linearly with K , so K cancels in \mathcal{I}_2 .

Proof. Step 1 (Second-order expansion). Expand $Y = F(\mathbf{X})$ around $\bar{\mathbf{x}} = c \mathbf{1}$. Let $\boldsymbol{\epsilon} = \mathbf{X} - c \mathbf{1}$. Then $F(\mathbf{X}) \approx c + Y_1 + Y_2$ where $Y_1 = \nabla F \cdot \boldsymbol{\epsilon} = (1/J) \mathbf{1} \cdot \boldsymbol{\epsilon} = \bar{\epsilon}$ and $Y_2 = \frac{1}{2} \boldsymbol{\epsilon}^T \nabla^2 F \boldsymbol{\epsilon}$.

Step 2 (Spectral decomposition). Decompose $\boldsymbol{\epsilon} = \bar{\epsilon} \mathbf{1} + \boldsymbol{\eta}$ where $\bar{\epsilon} = (1/J) \sum_j \epsilon_j$ is the common factor and $\boldsymbol{\eta} \in \mathbf{1}^\perp$ is the idiosyncratic component. Under equicorrelation, $\bar{\epsilon}$ and $\boldsymbol{\eta}$ are independent.

Step 3 (Channel separation). The linear term depends only on the common mode: $Y_1 = \bar{\epsilon}$.

The quadratic term depends only on the idiosyncratic modes. From Proposition 8.1:

$$Y_2 = -\frac{(1-\rho)}{2Jc} \|\boldsymbol{\eta}\|^2 = -\frac{K}{2(J-1)c} \|\boldsymbol{\eta}\|^2.$$

Step 4 (Diversity encoding). Under equicorrelation, the idiosyncratic modes Z_m (the projections of $\boldsymbol{\epsilon}$ onto $\mathbf{1}^\perp$) are i.i.d. with $\text{Var}[Z_m] = \tau^2(1-r)$, and $\|\boldsymbol{\eta}\|^2 = \sum_{m=2}^J Z_m^2$. Therefore $\mathbb{E}[\|\boldsymbol{\eta}\|^2] = (J-1)\tau^2(1-r)$, giving

$$\mathbb{E}[F(\mathbf{X})] = c + \mathbb{E}[Y_2] = c - \frac{K\tau^2(1-r)}{2c}, \quad (44)$$

which is (42). For the linear aggregate ($K=0$), $Y_2=0$ and $\mathbb{E}[F_{\text{lin}}] = c$.

Step 5 (Channel variances). The linear channel variance is

$$\text{Var}[Y_1] = \frac{\tau^2[1+r(J-1)]}{J}. \quad (45)$$

The quadratic channel variance uses $\text{Var}[\|\boldsymbol{\eta}\|^2] = 2(J-1)\tau^4(1-r)^2$ (from independence of the Z_m^2):

$$\text{Var}[Y_2] = \frac{K^2}{4(J-1)^2c^2} \cdot 2(J-1)\tau^4(1-r)^2 = \frac{K^2\tau^4(1-r)^2}{2(J-1)c^2}. \quad (46)$$

Step 6 (Fisher information). A shift $c \rightarrow c + \delta$ changes $\mathbb{E}[Y_1]$ by δ and $\mathbb{E}[Y_2]$ by $K\tau^2(1-r)\delta/(2c^2)$. Since the two channels are independent under equicorrelation, Fisher information is additive: $\mathcal{I}_i = (\partial_c \mathbb{E}[Y_i])^2 / \text{Var}[Y_i]$.

$$\mathcal{I}_1 = \frac{1}{\tau^2[1+r(J-1)]/J} = \frac{J}{\tau^2[1+r(J-1)]}, \quad (47)$$

$$\mathcal{I}_2 = \frac{K^2\tau^4(1-r)^2/(4c^4)}{K^2\tau^4(1-r)^2/[2(J-1)c^2]} = \frac{J-1}{2c^2}. \quad (48)$$

The K factors cancel in \mathcal{I}_2 , confirming the signal-to-noise ratio is curvature-invariant. The effective dimension $d_{\text{eff}} = \tau^2(\mathcal{I}_1 + \mathcal{I}_2)$ gives

$$d_{\text{eff}} = \frac{J}{1 + r(J - 1)} + \frac{(J - 1)\gamma^2}{2}. \quad (49)$$

At $K = 0$, $Y_2 = 0$ and only the first term survives: $d_{\text{eff}} = J/[1 + r(J - 1)]$. At $K > 0$, the quadratic channel adds $(J - 1)\gamma^2/2$ independent of the particular value of K . \square

Corollary 10.2 (Diversity detection threshold). *The diversity-encoding signal (42) exceeds the standard error of the CES output (i.e., is detectable in $O(1)$ observations) when*

$$K\gamma(1 - r) \gtrsim \sqrt{\frac{2[1 + r(J - 1)]}{J}}. \quad (50)$$

For sufficiently complementary inputs ($\rho < -1/(J - 1)$), $K > 1$ and the signal is detectable even for small γ .

Remark 10.3 (The role of K in correlation robustness). Unlike the superadditivity gap and strategic penalty—which tighten monotonically in K —the curvature parameter plays a qualitative rather than quantitative role in correlation robustness. $K > 0$ opens the quadratic channel that encodes idiosyncratic dispersion; $K = 0$ closes it. Conditional on the channel being open, its Fisher information is K -independent because signal and noise scale proportionally. The genuine K -dependent quantity is the expected-output sensitivity (42), which is first-order in K —the same scaling as superadditivity and strategic independence. All three curvature properties are $\Theta(K)$.

11. Strategic Independence

11.1 Setup

Consider J strategic agents, each controlling component $x_j \geq 0$. The aggregate $F(\mathbf{x})$ determines a common output. A coalition $S \subseteq [J]$ with $|S| = k$ can coordinate the levels $\{x_j\}_{j \in S}$.

Definition 11.1 (Manipulation gain). The **manipulation gain** of coalition S is

$$\Delta(S) = \sup_{\mathbf{x}_S \geq 0} \frac{v(S, \mathbf{x}_S) - v(S, \mathbf{x}_S^*)}{v(S, \mathbf{x}_S^*)}$$

where \mathbf{x}_S^* is the efficient allocation and $v(S, \mathbf{x}_S)$ is the coalition's Shapley value when playing \mathbf{x}_S against the efficient response of the other agents.

11.2 The Theorem

Theorem 11.2 (Strategic independence). *For all $\rho < 1$ and any coalition S with $|S| = k \leq J/2$:*

(i) $\Delta(S) \leq 0$. *No coalition can profitably manipulate the CES aggregate.*

(ii) *For any redistribution $\boldsymbol{\delta}_S$ with $\sum_{j \in S} \delta_j = 0$:*

$$\Delta(S) \leq -\frac{K}{2(J-1)} \cdot \frac{\|\boldsymbol{\delta}_S\|^2}{c^2} \leq 0. \quad (51)$$

The penalty tightens monotonically in K .

Proof. Step 1 (Qualitative). The characteristic function $v(S) = \max_{\mathbf{x}_S \geq 0} F(\mathbf{x}_S, \mathbf{0}_{-S})$ defines a convex cooperative game (Shapley, 1971), since F is concave. The Shapley value lies in the core; no deviation from the efficient allocation is profitable.

Step 2 (Quantitative). A coalition redistribution $\boldsymbol{\delta}_S$ with $\sum_{j \in S} \delta_j = 0$ changes output by $\Delta F = \frac{1}{2} \boldsymbol{\delta}_S^T H_{SS} \boldsymbol{\delta}_S + O(\|\boldsymbol{\delta}\|^3)$. From Proposition 8.1, for any $\boldsymbol{\delta}$ with $\sum_{j \in S} \delta_j = 0$:

$$\boldsymbol{\delta}_S^T H_{SS} \boldsymbol{\delta}_S = -\frac{K}{(J-1)c} \cdot \|\boldsymbol{\delta}_S\|^2.$$

The symmetric point is a strict local maximum of F over the coalition's feasible set. By symmetry, the efficient Shapley allocation assigns $v^*(S) = (k/J) \cdot c$. Since the CES game is convex, the coalition absorbs at least a (k/J) fraction of the output loss. Normalizing by $v^*(S) = (k/J) \cdot c$, the factors cancel:

$$\Delta(S) \leq -\frac{K}{2(J-1)} \cdot \frac{\|\boldsymbol{\delta}_S\|^2}{c^2} \leq 0. \quad \square$$

12. Network Scaling

The first three properties concern the *normalized* CES aggregate. The fourth concerns the *unnormalized* aggregate, which reveals how total value scales with network size.

Definition 12.1 (Unnormalized CES). The **unnormalized CES aggregate** with J components is

$$G(\mathbf{x}) = \left(\sum_{j=1}^J x_j^\rho \right)^{1/\rho}. \quad (52)$$

This equals $J^{1/\rho} \cdot F(\mathbf{x})$ when F uses equal weights $a_j = 1/J$.

Theorem 12.2 (Network scaling). *For J symmetric components with $x_j = c$ for all j :*

$$G(J) = J^{1/\rho} \cdot c. \quad (53)$$

The network scaling exponent is $1/\rho$:

- (i) $\rho = 1$ ($K = 0$): $G = Jc$. Linear scaling (Sarnoff's law).
- (ii) $\rho = 1/2$ ($K = (J - 1)/(2J)$): $G = J^2c$. Quadratic scaling (Metcalfe's law).
- (iii) $\rho < 1/2$ ($K > (J - 1)/(2J)$): $G = J^{1/\rho}c$ with $1/\rho > 2$. Super-Metcalfe scaling.

As $\rho \rightarrow 0$: scaling becomes exponential. For $\rho < 0$: $G \rightarrow 0$ as $J \rightarrow \infty$ —adding components dilutes value, because each additional complement creates a new potential bottleneck.

Proof. At the symmetric point: $G = (\sum_{j=1}^J c^\rho)^{1/\rho} = (J \cdot c^\rho)^{1/\rho} = J^{1/\rho} \cdot c$. \square

13. The Unified Theorem

Theorem 13.1 (CES Quadruple Role). Let F be a CES aggregate (1) with equal weights, $\rho < 1$, and $J \geq 2$. Define $K = (1 - \rho)(J - 1)/J$. Then $K > 0$, and:

- (a) **Superadditivity** (Theorem 9.1): $\text{gap} \geq \Omega(K) \cdot \text{diversity}$.
- (b) **Correlation robustness** (Theorem 10.1): $\mathbb{E}[F(\mathbf{X})] = c - \Theta(K) \cdot \tau^2(1-r)/(2c)$; curvature opens a quadratic channel closed under linear aggregation.
- (c) **Strategic independence** (Theorem 11.2): $\Delta(S) \leq -\Omega(K) \cdot \text{deviation}^2$.
- (d) **Network scaling** (Theorem 12.2): $G(J) = J^{1/\rho} \cdot c$, with exponent monotone in ρ .

Properties (a)–(c) are curvature effects: they arise from the Hessian eigenstructure at the symmetric point and tighten monotonically in K . Property (d) is an algebraic consequence of the power-sum form, governed by ρ directly rather than through the Hessian, but linked to (a)–(c) because K is monotone in ρ .

13.1 The Geometric Intuition

Properties (a)–(c) share a common origin: **the isoquant is not flat**. $\rho < 1$ is precisely the condition for non-flatness. $K = (1 - \rho)(J - 1)/J$ is precisely the degree of non-flatness. These three results are all consequences of the Hessian eigenstructure at the symmetric point (Proposition 8.1).

For **linear aggregation** ($\rho = 1$, $K = 0$): \mathcal{I}_1 is a hyperplane. Properties (a)–(c) vanish: $\text{gap} = 0$, diversity signal = 0, manipulation penalty = 0.

For **CES with** $\rho < 1$ ($K > 0$): \mathcal{I}_1 curves toward the origin. The curvature has three simultaneous consequences:

1. **Superadditivity.** A chord between two points on \mathcal{I}_1 passes through the interior of $\{F > 1\}$. Depth of penetration is $\Theta(K)$.
2. **Diversity encoding.** The expected CES output shifts by $\Theta(K)$ per unit of idiosyncratic variance. This encodes diversity information that linear aggregation destroys.

3. **Strategic stability.** Moving along \mathcal{I}_1 away from the balanced point always loses altitude at rate $\Theta(K)$.

Property (d) is different in character. Network scaling is a counting result about the *unnormalized* aggregate at the symmetric point: $G(J) = J^{1/\rho}c$ follows directly from the power-function form, not from isoquant curvature per se. Any homogeneous function of the form $(\sum x_j^\rho)^{1/\rho}$ exhibits this scaling, whether or not one interprets ρ as a curvature parameter. The connection to curvature is indirect: the same ρ that determines isoquant curvature also determines the scaling exponent, so $K > 0$ implies super-linear scaling as a logical consequence—but the mechanism is the power-sum structure, not the second-order geometry.

Property (d) is included in the unified theorem because the four properties are linked by the *parameter* ρ : any intervention that changes substitutability simultaneously changes all four. But the asymmetry should be noted: (a)–(c) are curvature effects (they arise from the Hessian and vanish when $K = 0$), while (d) is an algebraic effect that happens to share the same governing parameter.

13.2 Unified Scaling: All Three Are $\Theta(K)$

All three curvature properties scale linearly in K : (a) the superadditivity gap is $\Theta(K)$; (b) the diversity-encoding sensitivity is $\Theta(K)$; (c) the strategic penalty is $\Theta(K)$. This unity arises because each is a first-order consequence of the Hessian $\nabla^2 F = O(K)$. (The variance of the quadratic channel is $O(K^2)$, but both signal and noise inherit this scaling, so the Fisher information is K -independent; see Theorem 10.1.)

14. General Weights and the Secular Equation

With unequal weights $a_j > 0$ summing to 1, the symmetric point is replaced by the cost-minimizing point, the principal curvatures are no longer degenerate, and the curvature parameter acquires a weight-dispersion factor.

14.1 Effective Shares and the Cost-Minimizing Point

Define the **effective shares**

$$p_j = a_j^\sigma = a_j^{1/(1-\rho)}, \quad \Phi = \sum_{j=1}^J p_j, \quad (54)$$

and the **inverse effective shares** $w_j = 1/p_j = a_j^{-\sigma}$. For output level $c > 0$, the **cost-minimizing point** is

$$x_j^* = \frac{c p_j}{\Phi^{1/\rho}}, \quad j = 1, \dots, J. \quad (55)$$

14.2 The Hessian at General Weights

Proposition 14.1. *At the cost-minimizing point \mathbf{x}^* with general weights:*

$$\nabla^2 F|_{\mathbf{x}^*} = \frac{(1-\rho)g\Phi^{1/\rho}}{c\Phi} \left[\frac{\mathbf{p}\mathbf{p}^T}{\Phi} - \text{diag}(\mathbf{p}) \right] \quad (56)$$

where $g = \Phi^{(1-\rho)/\rho}$ is the common marginal product at \mathbf{x}^* .

14.3 The Secular Equation

Proposition 14.2 (Secular equation). *Let $w_j = 1/p_j = a_j^{-\sigma}$ be the ordered inverse shares with $w_{(1)} \leq w_{(2)} \leq \dots \leq w_{(J)}$. The principal curvatures of the CES isoquant at \mathbf{x}^* are determined by the constrained eigenvalues $\mu_1 < \mu_2 < \dots < \mu_{J-1}$ of $W = \text{diag}(w_1, \dots, w_J)$ restricted to $\mathbf{1}^\perp$, which satisfy the **secular equation***

$$\sum_{j=1}^J \frac{1}{w_j - \mu} = 0. \quad (57)$$

This equation has exactly $J - 1$ roots, one in each interval $(w_{(k)}, w_{(k+1)})$.

Remark 14.3 (Interlacing). The roots μ_k strictly interlace the poles $w_{(k)}$: $w_{(1)} < \mu_1 < w_{(2)} < \mu_2 < \dots < w_{(J-1)} < \mu_{J-1} < w_{(J)}$. This ensures all principal curvatures are positive for all $\rho < 1$ and all weight vectors.

14.4 The Generalized Curvature Parameter

Definition 14.4. The **generalized curvature parameter** is

$$K(\rho, \mathbf{a}) = (1-\rho) \frac{J-1}{J} \Phi^{1/\rho} R_{\min} \quad (58)$$

where $R_{\min} = \mu_1$ is the smallest root of the secular equation.

Proposition 14.5. *At equal weights ($a_j = 1/J$): $K(\rho, \mathbf{a})$ reduces to $(1-\rho)(J-1)/J$.*

Remark 14.6 (Equal-weight degeneracy). At equal weights, all inverse shares coincide ($w_j = J^\sigma$) and the secular equation degenerates—there are no intervals between distinct poles. By continuity, as $\mathbf{a} \rightarrow (1/J, \dots, 1/J)$, all $J - 1$ secular roots converge to $w = J^\sigma$, giving $R_{\min} = J^\sigma$. Then $\Phi^{1/\rho} R_{\min} = J^{-\sigma} \cdot J^\sigma = 1$, recovering $K = (1-\rho)(J-1)/J$.

14.5 General-Weight Quadruple Role

Theorem 14.7 (General-weight Quadruple Role). *Let F be a CES aggregate with weights \mathbf{a} and generalized curvature parameter $K(\rho, \mathbf{a})$. Then:*

(a) **Superadditivity.** The quantitative gap bound generalizes with $K(\rho, \mathbf{a})$ replacing K and κ_{\min} replacing κ^* .

(b) **Correlation robustness.** The diversity-encoding sensitivity generalizes with $K(\rho, \mathbf{a})$ replacing K . The secular roots determine how the idiosyncratic dispersion distributes across the $J - 1$ principal modes.

(c) **Strategic independence.** The manipulation penalty involves the **coalition curvature parameter**

$$K_S = (1 - \rho) \frac{k - 1}{k} \Phi_S^{1/\rho} R_{\min, S} \quad (59)$$

where $R_{\min, S}$ is the smallest root of the secular equation restricted to S . By the interlacing property, $K_S > 0$ for all coalitions of size $k \geq 2$.

(d) **Network scaling.** Define $\Psi = \sum_{j=1}^J a_j^{\sigma-1}$. The unnormalized aggregate at the cost-minimizing point satisfies $G = (\Psi/\Phi)^{1/\rho} \cdot c$, where $\Psi/\Phi = \mathbb{E}_p[1/a_j]$ is the effective-share-weighted average of $1/a_j$. At equal weights, $\Psi/\Phi = J$ and this reduces to $G = J^{1/\rho} c$.

15. Empirical Predictions

The emergence of CES generates five testable predictions, three of which are tested below.

P1 CES fit improves with aggregation: R_{CES}^2 rises at coarser levels; translog interaction terms shrink ($\|\beta\|/\|\alpha\| \rightarrow 0$).

P2 ρ is stable within levels: estimates at a given aggregation level are robust across methods and time periods.

P3 Tail exponents match σ : in sectors with elasticity $\hat{\sigma}$, the Pareto tail exponent satisfies $\hat{\zeta} \approx \hat{\sigma}$.

P4 Translog coefficients shrink monotonically: the ratio $\|\beta\|/\|\alpha\|$ decreases with aggregation level—the direct test of transient terms contracting.

P5 Complementary sectors have heavier tails: $\partial \hat{\zeta} / \partial \hat{\rho} > 0$ cross-sectionally, provided $T > 0$.

P3 and P5 require firm-level size distributions by sector and are left for future work. P1, P2, and P4 are tested using Monte Carlo simulation, manufacturing production data, and industrial production indices.

15.1 Monte Carlo: CES Emerges from Non-CES Micro Data

The emergence prediction is first verified computationally. The simulation uses 200 firms with heterogeneous decreasing-returns Cobb-Douglas production functions ($y_j = A_j k_j^{\alpha_j \nu} l_j^{(1-\alpha_j)\nu}$, $\nu = 0.85$, $\alpha_j \sim \text{Beta}(2, 3)$) that profit-maximize at common factor prices (r_t, w_t) following AR(1) processes over $T = 500$ periods. Firms are aggregated hierarchically by summation: $200 \rightarrow 40 \rightarrow 8 \rightarrow 2 \rightarrow 1$. At each level, CES and translog are estimated on the aggregate (Y, K, L) time series and averaged across 30 Monte Carlo

replications.

CES achieves $R^2 = 0.97$ at all aggregate levels (Table 1), confirming the [Houthakker \(1955\)](#) mechanism: heterogeneous firms with *different* production functions generate CES at the aggregate, exactly as the emergence theorem predicts. The estimated $\hat{\rho} \approx 0$ reflects the Cobb-Douglas micro structure.

In a separate simulation, 120 firms with heterogeneous *translog* production functions (interaction scale = 0.15) are aggregated with common-factor inputs. The translog interaction ratio $\|\beta\|/\|\alpha\|$ decays monotonically from 0.35 at the first aggregate level to 0.16 at the coarsest—a 53% reduction—directly confirming that non-CES terms are transient terms that shrink under aggregation.

15.2 NBER-CES Manufacturing Database

The second test uses the NBER-CES Manufacturing Industry Database ([Becker, Gray, and Marvakov, 2021](#)): value added, capital stock, and production worker hours for 364 six-digit NAICS industries, 1977–2018. Aggregating from 6-digit to 2-digit yields five hierarchy levels (Table 2).

Both key predictions are confirmed with monotone patterns across all five levels:

- **CES fit improves:** Median R_{CES}^2 rises from 0.944 (6-digit) to 0.975 (2-digit).
- **Translog terms shrink:** The interaction ratio decays monotonically from 0.137 to 0.068—a 50% reduction.

The median curvature is $\tilde{\rho} \approx -2.7$ ($\sigma \approx 0.27$), strongly complementary. This is below the consensus range $\sigma \in [0.4, 0.7]$ established by [Chirinko \(2008\)](#)’s survey of the estimation literature and the structural estimates of [Antràs \(2004\)](#) and [León-Ledesma, McAdam, and Willman \(2010\)](#). The discrepancy is informative: [Antràs \(2004\)](#) estimates σ from cost shares in a supply-side system, while [León-Ledesma, McAdam, and Willman \(2010\)](#) show that joint estimation of the CES with factor-augmenting technical change requires careful normalization to avoid biased $\hat{\sigma}$. The time-series identification used here captures within-industry substitution, while cross-sectional approaches (including [Oberfield and Raval 2014](#)) also reflect between-plant reallocation, which raises apparent substitutability—exactly the pattern the stability theorem predicts: aggregation over heterogeneous factor ratios mimics substitution even when micro-level complementarity is strong. The interquartile range of $\hat{\rho}$ narrows from 1.6 (6-digit) to 0.07 (2-digit), confirming that aggregation concentrates the curvature distribution (P2).

15.3 FRED Industrial Production Hierarchy

For a complementary test on high-frequency data, the FRED Industrial Production index hierarchy is used (monthly, 1972–2025, $T = 649$), which provides six aggregation nodes with $J \in \{2, \dots, 9\}$ sub-components each. Unlike the NBER-CES data (which estimates

production functions), the FRED test estimates how sub-indices combine into aggregates.

At every node, CES achieves $R^2 > 0.998$ (Table 3). The overidentification test uses $\hat{\rho}_{\text{NLS}}$ at each node to predict the common-factor R^2 (variance explained by the first principal component):

$$\chi^2 = \sum_{\ell=1}^6 \frac{(R_{\text{pred},\ell}^2 - R_{\text{obs},\ell}^2)^2}{\widehat{\text{Var}}(R_{\text{obs},\ell}^2)} = 6.37, \quad df = 5, \quad p = 0.27. \quad (60)$$

The null that a single CES parameter per node generates both the production-function fit and the covariance structure cannot be rejected. This is the quadruple role theorem (Theorem 13.1) in action: ρ simultaneously governs production complementarity and statistical correlation.

Table 1: Monte Carlo: CES emergence under hierarchical aggregation

Simulation	Level	n	$\ \beta\ /\ \alpha\ $	R_{CES}^2	% F -rej.	$\hat{\rho}$
Houthakker	1	40	0.106	0.973	80%	−0.05
	2	8	0.175	0.973	96%	0.02
	3	2	0.229	0.973	100%	0.08
Translog decay	1	24	0.351	0.555	68%	0.14
	2	6	0.294	0.487	73%	0.30
	3	2	0.173	0.431	90%	0.57

Notes: 30 MC replications, $T = 500$. Houthakker: 200 DRS Cobb-Douglas firms ($\nu = 0.85$, $\alpha_j \sim \text{Beta}(2, 3)$) at common factor prices. Translog decay: 120 firms with heterogeneous translog (interaction scale = 0.15). Aggregation by summation. $\|\beta\|/\|\alpha\|$: translog interaction ratio.

Table 2: CES emergence in NBER-CES Manufacturing data under NAICS aggregation

NAICS	n	R_{CES}^2	R_{TL}^2	$\ \beta\ /\ \alpha\ $	$\tilde{\rho}$	IQR(ρ)
6-digit	364	0.944	0.970	0.137	−2.70	1.63
5-digit	180	0.950	0.977	0.125	−2.74	1.49
4-digit	86	0.953	0.981	0.112	−2.75	1.10
3-digit	21	0.971	0.990	0.087	−2.87	1.13
2-digit	3	0.975	0.993	0.068	−2.49	0.07

Notes: NBER-CES Manufacturing Industry Database (1977–2018, 2012 NAICS). Value added regressed on capital stock and production hours with time trend. Medians across industries at each level. $\tilde{\rho}$: median CES curvature. IQR: interquartile range.

16. Extensions and Discussion

16.1 Nested CES and Multi-Scale Structure

A hierarchical CES structure with different ρ at each level corresponds to a non-trivial trajectory through ρ -space: the system has not converged to a single fixed point. The

Table 3: CES fit in FRED Industrial Production hierarchy

Aggregate	Depth	J	$\hat{\rho}$	95% CI	R_{CES}^2	$\ \beta\ /\ \alpha\ $	R_{common}^2
Total IP	0	3	-0.073	$[-0.13, -0.01]$	0.9999	0.176	0.492
Manufacturing	1	2	-0.299	$[-0.31, -0.29]$	1.0000	0.188	0.821
Nondurable Goods	2	9	0.990	$[0.89, 1.09]$	0.9997	0.236	0.407
Durable Goods	2	8	0.152	$[0.14, 0.17]$	0.9992	0.693	0.504
Computer/Elec.	3	5	-0.191	$[-0.20, -0.18]$	0.9999	0.179	0.341
Transport Equip.	3	2	0.673	$[0.60, 0.75]$	0.9984	0.277	0.651

Notes: $\hat{\rho}$ by NLS; 95% CI from profile likelihood. Overidentification: $\chi^2 = 6.37$ ($df = 5$, $p = 0.27$). Data: FRED IP indices, monthly, 1972–2025.

β -function $d\rho/d\log s = \beta(\rho)$ encodes how effective complementarity changes with scale. The fixed points are: $\rho = 1$ (always stable from above), $\rho = 0$ (stable under mild conditions), $\rho \rightarrow -\infty$ (unstable).

This structure connects to the two-level CES of Sato (1967), which nests CES sub-aggregates within a CES outer aggregate, allowing different elasticities at different levels. The emergence framework provides a theoretical justification for Sato’s construction: if scale consistency holds *within* each level but the curvature parameter varies *across* levels, the result is exactly a nested CES hierarchy. The emergence theorem (Theorem 3.2) operates level by level, selecting a (potentially different) power mean at each scale.

The empirical elasticity ($\sigma \approx 0.6$ – 0.8 , implying $\rho \approx -0.25$ to -0.67) sits between Cobb-Douglas and Leontief, consistent with slow convergence toward $\rho = 0$.

16.2 Weighted CES and Asymmetric Inputs

The emergence theorem assumes symmetric inputs. Corollary 16.1 shows share asymmetry (different w_j) is compatible with emergence. But *elasticity asymmetry* (different ρ_{jk} for different input pairs) breaks the result. However, the aggregation operator still averages ρ_{jk} within blocks, yielding *local* emergence.

Corollary 16.1 (Weighted emergence). *Let $F : \mathbb{R}_+^J \rightarrow \mathbb{R}_+$ be continuous, strictly increasing, homogeneous of degree one, and scale-consistent under a weighted partition. Then F is weighted CES with weights proportional to the atoms of the partition.*

16.3 The Translog as an Irrelevant Perturbation

The translog production function adds transient terms to the Cobb-Douglas fixed point: the β_{jk} terms break scale consistency and vanish under aggregation. 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: why CES fits aggregate data well despite being “more restrictive” than translog. The translog’s extra flexibility is precisely the kind that doesn’t survive aggregation. Estimating a translog at the macro level is fitting noise.

Jones (2005) posed the “shape of production functions” question sharply: given that growth outcomes depend critically on whether $\sigma \gtrless 1$, what determines the functional form? His analysis showed that the direction of technical change and the shape of the production function are jointly determined, but took the CES form as given. The emergence results provide the missing foundation: the shape is not a free parameter but is forced by the multi-scale structure of the economy. The relevant question is not “CES or translog?” but “what determines ρ ?”

16.4 Tightness of the Quadruple Role Bounds

All bounds become equalities in limit cases: (a) gap vanishes for proportional inputs; (b) diversity signal vanishes as $K \rightarrow 0$ or $r \rightarrow 1$; (c) penalty vanishes as $K \rightarrow 0$ or $k/J \rightarrow 0$. For $\rho < 0$, the curvature increases away from the symmetric point, so local bounds are conservative.

16.5 What This Paper Does Not Cover

This paper proves the emergence of CES and the static properties of a single CES aggregate. It does not address:

- *Dynamics.* How the curvature parameter governs transition dynamics, conservation laws, and business cycles.
- *Endogenous ρ .* How ρ itself evolves through learning, competition, and institutional feedback.
- *Information friction.* The (ρ, T) regime diagram and the role of the information friction parameter T .
- *Multiple hierarchical levels.* How curvature interacts across a multi-level economy with timescale separation.

17. 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 converge on this conclusion, and their convergence is itself a mathematical fact (Theorem 6.1).

The forced functional form then does four things simultaneously, all controlled by a single dimensionless parameter $K = (1 - \rho)(J - 1)/J$: it creates value from diversity (superadditivity), extracts signal from noise (correlation robustness), resists strategic manipulation (strategic independence), and determines network scaling. Properties (a)–(c)

are three views of the curvature of CES isoquants; property (d) extends this to endogenous network size. When $K = 0$, all four vanish. The parameter ρ simultaneously serves as the aggregation-invariant exponent, the Rényi entropy order, the Tsallis deformation parameter, and the aggregation-invariant class label.

In the language of aggregation theory: 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.
- Antràs, Pol. 2004. “Is the U.S. Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution.” *Contributions to Macroeconomics* 4(1): Article 4.
- 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.
- Becker, Randy, Wayne Gray, and Jordan Marvakov. 2021. “NBER-CES Manufacturing Industry Database (1958–2018, Technical Notes).” National Bureau of Economic Research.
- Blackorby, Charles, and William Schworm. 1993. “The Implications of Additive Structure in the Theory of Consistent Aggregation.” In *Contributions to Mathematical Economics and Optimization Theory*, ed. Wolfgang Eichhorn, 197–215. Berlin: Springer.
- Chirinko, Robert S. 2008. “ σ : The Long and Short of It.” *Journal of Macroeconomics* 30(2): 671–686.
- Christensen, Laurits R., Dale W. Jorgenson, and Lawrence J. Lau. 1973. “Transcendental Logarithmic Production Frontiers.” *Review of Economics and Statistics* 55(1): 28–45.
- de La Grandville, Olivier. 1989. “In Quest of the Slutsky Diamond.” *American Economic Review* 79(3): 468–481.
- Diamond, Peter A., and Daniel McFadden. 1974. “Some Uses of the Expenditure Function in Public Finance.” *Journal of Public Economics* 3(1): 3–21.
- Dixit, Avinash K., and Joseph E. Stiglitz. 1977. “Monopolistic Competition and Optimum Product Diversity.” *American Economic Review* 67(3): 297–308.
- do Carmo, Manfredo P. 1992. *Riemannian Geometry*. Boston: Birkhäuser.
- Gabaix, Xavier. 2009. “Power Laws in Economics and Finance.” *Annual Review of Economics* 1(1): 255–294.
- Gorman, W. M. 1968. “Measuring the Quantities of Fixed Factors.” In *Value, Capital, and Growth: Papers in Honour of Sir John Hicks*, ed. J. N. Wolfe, 141–172. Edinburgh: Edinburgh University Press.

- Hardy, G. H., J. E. Littlewood, and G. Pólya. 1952. *Inequalities*. 2nd ed. Cambridge: Cambridge University Press.
- Houthakker, Hendrik S. 1955. “The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis.” *Review of Economic Studies* 23: 27–31.
- Jones, Charles I. 2005. “The Shape of Production Functions and the Direction of Technical Change.” *Quarterly Journal of Economics* 120(2): 517–549.
- Kmenta, Jan. 1967. “On Estimation of the CES Production Function.” *International Economic Review* 8(2): 180–189.
- Klump, Rainer, and Olivier de La Grandville. 2000. “Economic Growth and the Elasticity of Substitution: Two Theorems and Some Suggestions.” *American Economic Review* 90(1): 282–291.
- Kolmogorov, Andrey N. 1930. “Sur la Notion de la Moyenne.” *Atti della Reale Accademia Nazionale dei Lincei. Rendiconti* 12: 388–391.
- León-Ledesma, Miguel A., Peter McAdam, and Alpo Willman. 2010. “Identifying the Elasticity of Substitution with Biased Technical Change.” *American Economic Review* 100(4): 1330–1357.
- 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.
- Shapley, Lloyd S. 1971. “Cores of Convex Games.” *International Journal of Game Theory* 1: 11–26.
- Tsallis, Constantino. 1988. “Possible Generalization of Boltzmann-Gibbs Statistics.” *Journal of Statistical Physics* 52(1–2): 479–487.