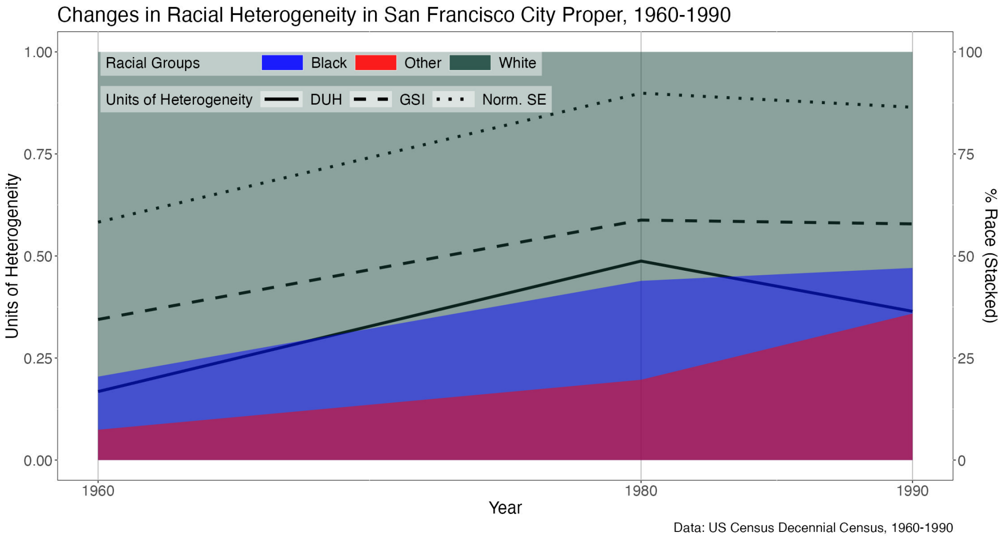
Descriptive Units of Heterogeneity: An Axiomatic Approach to Measuring Heterogeneity

Willy Chen

AFRE Brown Bag

MICHIGAN STATE UNIVERSITY

Empirical Motivation



What is Heterogeneity in a System?

A system $s = (s_1, s_2, \ldots)$ is an ordered tuple of non-negative real numbers that is not the null tuple. Denote the set of all such systems $\mathcal{S}$.

What is Heterogeneity in a System?

A system $s = (s_1, s_2, \ldots)$ is an ordered tuple of non-negative real numbers that is not the null tuple. Denote the set of all such systems $\mathcal{S}$.

$\forall s \in \mathcal{S}$, denote

- The length of s as $|s|$.
- The "total population" of s as $||s||_1 = \sum_{g=1}^{|s|} s_g$.
- The mean group size of s as $\mu(s) = \frac{||s||_1}{|s|}$.

What is Heterogeneity in a System?

A system $s = (s_1, s_2, \ldots)$ is an ordered tuple of non-negative real numbers that is not the null tuple. Denote the set of all such systems $\mathscr{S}$.

$\forall s \in \mathscr{S}$, denote

- The length of s as $|s|$.
- The "total population" of s as $||s||_1 = \sum_{g=1}^{|s|} s_g$.
- The mean group size of s as $\mu(s) = \frac{||s||_1}{|s|}$.

A function $\Phi : \mathscr{S} \to \mathbb{R}$ is a measure of heterogeneity if, $\forall s, s' \in \mathscr{S}$,

$$\Phi(s) \geq \Phi(s') \iff s \text{ is weakly more heterogeneous } (\succsim) \text{ than } s'.$$

Defining Maximum and Minimum Heterogeneity

Heterogeneity is defined by the abundant variation in group sizes, and homogeneity the lack thereof.

Defining Maximum and Minimum Heterogeneity

Heterogeneity is defined by the abundant variation in group sizes, and homogeneity the lack thereof.

For example, $(50, 50)$ vs. $(100, 0)$.

Defining Maximum and Minimum Heterogeneity

Heterogeneity is defined by the abundant variation in group sizes, and homogeneity the lack thereof.

For example, $(50, 50)$ vs. $(100, 0)$.

A system $\bar{s}$ is maximally heterogeneous if:

$$\exists k \in \mathbb{R}_{++},\ \bar{s} = k \cdot (1, 1, \ldots, 1).$$

Defining Maximum and Minimum Heterogeneity

Heterogeneity is defined by the abundant variation in group sizes, and homogeneity the lack thereof.

For example, $(50, 50)$ vs. $(100, 0)$.

A system $\bar{s}$ is maximally heterogeneous if:

$$\exists k \in \mathbb{R}_{++},\ \bar{s} = k \cdot (1, 1, \ldots, 1).$$

A system $\underline{s}$ is minimally heterogeneous (or perfectly homogeneous) if all but one element of $\underline{s}$ are 0.

Repurposed Measures of Heterogeneity

Repurposed Measures of Heterogeneity

- Measures of dispersion/inequality such as Gini coefficient, Atkinson index, Hoover index, etc.

Repurposed Measures of Heterogeneity

- Measures of dispersion/inequality such as Gini coefficient, Atkinson index, Hoover index, etc.
- Measures of concentration such as Herfindahl-Hirschman index, Shannon's entropy, etc.

Repurposed Measures of Heterogeneity

- Measures of dispersion/inequality such as Gini coefficient, Atkinson index, Hoover index, etc.
- Measures of concentration such as Herfindahl-Hirschman index, Shannon's entropy, etc.

Problem: If we treat heterogeneity as a distributional property without value-judgments, the there are cases when dispersion and deconcentration are sufficient.

Measures of Inequality

Measures of inequality quantify heterogeneity by the distance between observed distribution and the uniform distribution.

Measures of Inequality

Measures of inequality quantify heterogeneity by the distance between observed distribution and the uniform distribution.

- Requires distributional assumptions (Gini).

Measures of Inequality

Measures of inequality quantify heterogeneity by the distance between observed distribution and the uniform distribution.

- Requires distributional assumptions (Gini).
- Require arbitrary normative judgments (Atkinson and Genralized Gini).

Measures of Inequality

Measures of inequality quantify heterogeneity by the distance between observed distribution and the uniform distribution.

- Requires distributional assumptions (Gini).
- Require arbitrary normative judgments (Atkinson and Genralized Gini).
- Require implicit order of outcomes (Atkinson, Generalized Gini, and Gini).

Measures of Inequality

Measures of inequality quantify heterogeneity by the distance between observed distribution and the uniform distribution.

- Requires distributional assumptions (Gini).
- Require arbitrary normative judgments (Atkinson and Genralized Gini).
- Require implicit order of outcomes (Atkinson, Generalized Gini, and Gini).
- Lacks compressibility and the sensitivity to transfers (Hoover).

Measures of Concentration

Measures of concentration quantify heterogeneity by the non-dominance of one or a few groups in a system (deconcentration).

Measures of Concentration

Measures of concentration quantify heterogeneity by the non-dominance of one or a few groups in a system (deconcentration).

- Not sensitive to redistribution between small groups when one group is sufficiently large ((HHI and SE) .
- Discard information provided by the presence of groups with 0 population (HHI and SE).

Descriptive Units of Heterogeneity

Let σ of s be the permutation of s such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{|s|}$. σ is called the *ordered system* of s.

Define

$$\hat{\sigma}_1 = \frac{\sigma_1}{||s||_1} \text{ and } \tilde{\sigma}_g = \frac{\sigma_g}{||s||_1 - \sigma_1}$$

The Descriptive Units of Heterogeneity (DUH) is defined as:

$$DUH_p(s) = \frac{ln(\hat{\sigma}_1)}{ln(|s|)} \cdot \left[\left(\sum_{g=2}^{|s|} \left| \tilde{\sigma}_g - \frac{1}{|s| - 1} \right|^p \right)^{\frac{1}{p}} - 1 \right]$$

Fundamental Axioms

Axioms that induce partial ordering by pinning down

1. When two systems of $|s|$ groups are equally heterogeneous
2. How to order the heterogeneity of two systems of $|s|$ groups that are marginally different

[GSYM] Group Symmetry

For any permutation $\pi(s)$ of s, $\Phi(s) = \Phi(\pi(s))$.

For example, take $s, s', s'' \in \mathcal{S}$,

$$\begin{array}{c} s \\ \begin{array}{|c|}\hline s_1 \\ \hline s_2 \\ \hline s_3 \\ \hline \end{array} \end{array} \sim \begin{array}{c} s' \\ \begin{array}{|c|}\hline s_2 \\ \hline s_1 \\ \hline s_3 \\ \hline \end{array} \end{array} \sim \begin{array}{c} s'' \\ \begin{array}{|c|}\hline s_3 \\ \hline s_2 \\ \hline s_1 \\ \hline \end{array} \end{array}$$

[INV] Scale Invariance

A measure of heterogeneity Φ satisfies the property of Scale Invariance if for any system s and a scalar $\lambda \in \mathbb{R}_{++}$, $\Phi(s) = \Phi\left(\lambda \cdot s\right)$.

For example, take $s, s' \in \mathcal{S}$, $\lambda \in \mathbb{R}_{++}$,

$$\begin{array}{ccc} s & & s' \\ \begin{array}{|c|}\hline s_1 \\ \hline s_2 \\ \hline s_3 \\ \hline \end{array} & \sim & \begin{array}{|c|}\hline \lambda s_1 \\ \hline \lambda s_2 \\ \hline \lambda s_3 \\ \hline \end{array} \end{array}$$

[PT] Principle of Transfers

Let σ be the ordered system of s. Let $e_i^{|s|}$ be an ordered tuple of length $|s|$ such that its i^{th} element is some ε and the rest are 0. A measure of heterogeneity Φ satisfies the Principle of Transfers if $\forall i < j \leq |\sigma|$ and $\varepsilon \in \mathbb{R}_+$,

$$\begin{cases} \sigma_i - \sigma_j \geq 2\varepsilon \\ \sigma_i - \sigma_{i+1} \geq \varepsilon \\ \sigma_{j-1} - \sigma_j \geq \varepsilon \end{cases} \quad \text{together imply } \Phi(\sigma) < \Phi\left(\sigma - e_i^{|s|} + e_j^{|s|}\right).$$

[PT] Principle of Transfers

Let σ be the ordered system of s. Let $e_i^{|s|}$ be an ordered tuple of length $|s|$ such that its i^{th} element is some ε and the rest are 0. A measure of heterogeneity Φ satisfies the Principle of Transfers if $\forall i < j \leq |\sigma|$ and $\varepsilon \in \mathbb{R}_+$,

$$\begin{cases} \sigma_i - \sigma_j \geq 2\varepsilon \\ \sigma_i - \sigma_{i+1} \geq \varepsilon \\ \sigma_{j-1} - \sigma_j \geq \varepsilon \end{cases} \quad \text{together imply } \Phi(\sigma) < \Phi\left(\sigma - e_i^{|s|} + e_j^{|s|}\right).$$

$$\begin{array}{c} \sigma \\ \begin{array}{|c|} \hline \sigma_1 \\ \hline \sigma_2 \\ \hline \sigma_3 \\ \hline \end{array} \end{array} \prec \begin{array}{c} \sigma' \\ \begin{array}{|c|} \hline \sigma_1 - \varepsilon \\ \hline \sigma_2 + \varepsilon \\ \hline \sigma_3 \\ \hline \end{array} \end{array} \prec \begin{array}{c} \sigma'' \\ \begin{array}{|c|} \hline \sigma_1 - \varepsilon \\ \hline \sigma_2 \\ \hline \sigma_3 + \varepsilon \\ \hline \end{array} \end{array}, \quad \text{i.e. } \Phi(\sigma) < \Phi(\sigma') < \Phi(\sigma''),$$

A Different Sufficient Condition for GSYM

- Existing units (satisfying INV) has the same arithmetic treatment for every group in the system, partially to satisfy GSYM

A Different Sufficient Condition for GSYM

- Existing units (satisfying INV) has the same arithmetic treatment for every group in the system, partially to satisfy GSYM
  - For example, HHI is the sum of squares of every group's proportion, instead of sum of squares of some groups' proportions and cubes of other groups' proportions.

A Different Sufficient Condition for GSYM

- Existing units (satisfying INV) has the same arithmetic treatment for every group in the system, partially to satisfy GSYM
  - For example, HHI is the sum of squares of every group's proportion, instead of sum of squares of some groups' proportions and cubes of other groups' proportions.
- This uniform treatment can be relaxed by treating the same type of groups the same.

What I propose is to treat the largest group separate from the rest of the groups.

What Matters in Measuring Heterogeneity?

The heterogeneity of system S is measured by $\Phi : S \to \mathbb{R}$ such that

$$\Phi(s) \geq \Phi(s') \iff s \text{ is weakly more heterogeneous than } s'.$$

What Matters in Measuring Heterogeneity?

The heterogeneity of system S is measured by $\Phi : S \to \mathbb{R}$ such that

$$\Phi(s) \geq \Phi(s') \iff s \text{ is weakly more heterogeneous than } s'.$$

- Relative size of the largest group: $\hat{\sigma}_1 = \frac{\sigma_1}{||s||_1}$

What Matters in Measuring Heterogeneity?

The heterogeneity of system S is measured by $\Phi : S \to \mathbb{R}$ such that

$$\Phi(s) \geq \Phi(s') \iff s \text{ is weakly more heterogeneous than } s'.$$

- Relative size of the largest group: $\hat{\sigma}_1 = \frac{\sigma_1}{||s||_1}$
- Relative size(s) of the minority group(s): $\tilde{\sigma}_g = \frac{\sigma_g}{||s||_1 - \sigma_1}$, $g \in \{2, \ldots, |s|\}$

A unit of heterogeneity can be thought of as $\Phi = \Phi(\phi, \psi)$ where
$\phi(\hat{\sigma}_1)$ is the influence of the relative size of the largest group
$\psi(\tilde{\sigma}_2, \ldots, \tilde{\sigma}_G)$ the influence of the relative sizes of the minority group.

Characterization Axioms

Axioms that can be combined to uniquely characterize measures with total ordering and cardinal interpretation.

[IND] Independence

Let σ be the ordered system of s. A measure of heterogeneity $\Phi(s)$ satisfies Independence if it is a composite function of $\phi : \mathcal{S} \to \mathbb{R}$ and $\psi : \mathcal{S} \to \mathbb{R}$ such that $\psi\left(s\right) = \psi\left(c, \sigma_2, \ldots, \sigma_{|s|}\right), \forall c \in \mathbb{R}_{++}$.

Using Evenness for ψ

I use the L^p-norm between the minority distribution and the uniform distribution to quantify evenness in minority distribution.

Using Evenness for ψ

I use the L^p-norm between the minority distribution and the uniform distribution to quantify evenness in minority distribution.

Definition: Take $p \in [1, \infty)$. The function $\psi_p : \mathcal{S} \to \mathbb{R}$ defined as:

$$\psi_p(s) = 1 - \left(\sum_{g=2}^{|s|} \left| \tilde{\sigma}_g - \frac{1}{|s| - 1} \right|^p \right)^{\frac{1}{p}},$$

is a measure of evenness in the distribution of minority groups.

Proposition

Let s be an arbitrary system such that $|s| \geq 4$. Let Φ be a measure of heterogeneity that satisfies GSYM, INV, and IND. Holding $\hat{\sigma}_1$ constant, if Φ is strictly increasing in ψ_p, then Φ satisfies PT if and only if $p > 1$.

[PPT] Principle of Proportional Transfers

Let σ be the ordered system of s. Let $e_i^{|s|}$ be an ordered tuple of length $|s|$ such that its i^{th} element is some ε and the rest are 0. A measure of heterogeneity $\Phi(s)$ satisfies the Principle of Proportional Transfers if $\forall \varepsilon \in \mathbb{R}_+, \exists \alpha \in \mathbb{R}_{++}$

$$\begin{cases} \frac{\sigma_1 - \varepsilon}{||s||_1} = \left(\frac{\sigma_1}{||s||_1}\right)^\alpha \\ \sigma_1 - \varepsilon \geq \sigma_2 + \tilde{\sigma}_2 \cdot \varepsilon \end{cases} \quad \text{together imply } \Phi\left(\sigma - e_1^{|s|} + \sum_{g=2}^{|s|} \tilde{\sigma}_g \cdot e_g^{|s|}\right) = \alpha \cdot \Phi\left(\sigma\right).$$

In other words, *holding the order of groups constant*, a transfer from the largest group proportionally to the minority groups that reduces $\hat{\sigma}_1$ to $(\hat{\sigma}_1)^\alpha$ increases heterogeneity by a factor of α.

PPT Example

$$
\begin{array}{ccc}
\sigma & & \sigma' \\
\boxed{\begin{array}{c} \tilde{\sigma}_1 \\ \hline \hat{\sigma}_2 \\ \hline \hat{\sigma}_3 \end{array}} & \prec & \boxed{\begin{array}{c} \hat{\sigma}_1^{\alpha} \\ \hline \hat{\sigma}_2 + \frac{\hat{\sigma}_2}{\hat{\sigma}_2+\hat{\sigma}_3}\left(\hat{\sigma}_1 - \hat{\sigma}_1^{\alpha}\right) \\ \hline \hat{\sigma}_3 + \frac{\hat{\sigma}_3}{\hat{\sigma}_2+\hat{\sigma}_3}\left(\hat{\sigma}_1 - \hat{\sigma}_1^{\alpha}\right) \end{array}}
\end{array}.
$$

A measure Φ satisfying GSYM, INV, and PPT would yield:

$$\Phi\left(\sigma\right) < \alpha\Phi\left(\sigma\right) = \Phi\left(\sigma'\right).$$

[CON] Contractibility

Let s be an arbitrary system. Let s' be the concatenation of s and the tuple (0) such that $s' = (s, 0)$. Let σ and σ' denote the ordered systems of s and s'. A measure of heterogeneity Φ satisfies Contractibility if

$$\sigma_2 > 0 \ \Rightarrow \ \Phi\left(\sigma'\right) < \Phi\left(\sigma\right).$$

[UNI] Unity

Let σ be the ordered system of s. A measure of heterogeneity Φ satisfies Unity if

$$\Phi\left(s\right) = 0 \iff \hat{\sigma}_1 = 1$$

and

$$\Phi\left(s\right) = 1 \iff \hat{\sigma}_1 = \hat{\sigma}_2 = \cdots = \hat{\sigma}_{|s|} = \frac{1}{|s|}$$

Descriptive Units of Heterogeneity

Let σ be the ordered system of s. Denote $\hat{\sigma}_1 = \frac{\sigma_1}{||s||_1}$ and $\tilde{\sigma}_g = \frac{\sigma_g}{||s||_1 - \sigma_1}$, where $g \in \{2, \ldots, |s|\}$ and $p \in (1, \infty)$. The family of descriptive units of heterogeneity (DUH) of the system s with $|s| \geq 2$ is:

$$DUH_p\left(s\right) = \frac{ln\left(\hat{\sigma}_1\right)}{ln(|s|)} \cdot \left[\left(\sum_{g=2}^{|s|} \left|\tilde{\sigma}_g - \frac{1}{|s| - 1}\right|^p\right)^{\frac{1}{p}} - 1\right].$$

Descriptive Units of Heterogeneity

Let σ be the ordered system of s. Denote $\hat{\sigma}_1 = \frac{\sigma_1}{||s||_1}$ and $\tilde{\sigma}_g = \frac{\sigma_g}{||s||_1 - \sigma_1}$, where $g \in \{2, \ldots, |s|\}$ and $p \in (1, \infty)$. The family of descriptive units of heterogeneity (DUH) of the system s with $|s| \geq 2$ is:

$$DUH_p(s) = \frac{ln(\hat{\sigma}_1)}{ln(|s|)} \cdot \left[\left(\sum_{g=2}^{|s|} \left| \tilde{\sigma}_g - \frac{1}{|s|-1} \right|^p \right)^{\frac{1}{p}} - 1 \right].$$

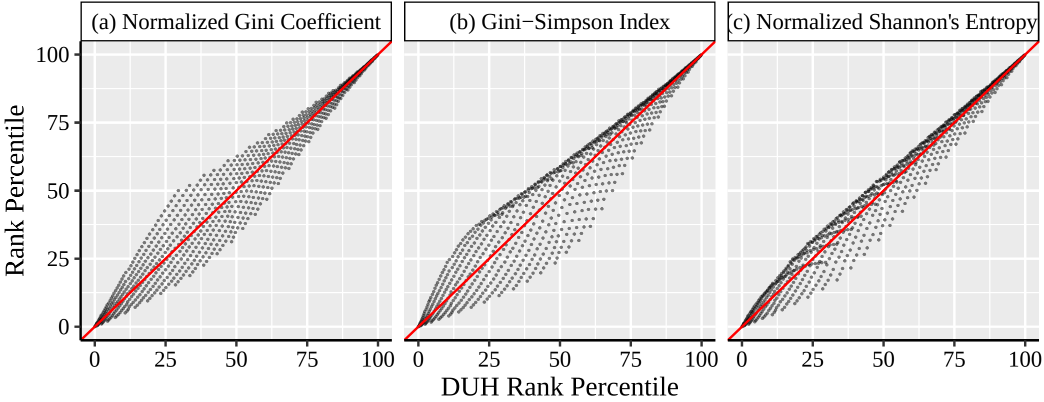
Theorem:
The descriptive units of heterogeneity constitute a uniquely determined family of measures—up to positive scalar multiplication—that incorporate evenness through the function ψ_p, and satisfy GSYM, INV, PT, IND, PPT, CON, and UNI.

Existing Measures Satisfy Some Axioms

Type	Axiom	Gini	DUH	HHI	SE
Fundamental	Type Symmetry	✓	✓	✓	✓
	Scale Invariance	✓	✓	✓	✓
	Principle of Transfers	✓	✓	✓	✓
Characterization	Independence	×	✓	✓	✓
	Principle of Proportional Transfers	×	✓	×	×
	Contractibility	✓	✓	×	×
	Unity	×	✓	×	×
	Expandability	×	×	✓	✓
	Replication Principle	×	×	✓	×
	Shannon's Additivity	×	×	×	✓

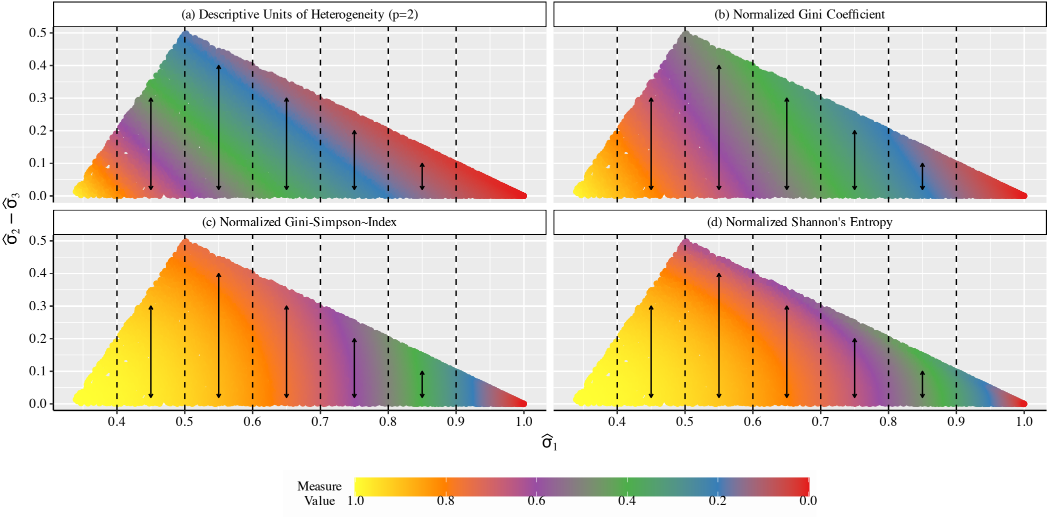
Comparing Measures

Figure 1: Rank Correlation between Measures over Systems with $|\sigma| = 3$ and $||\sigma||_1 = 100$



Comparing Measures

Figure 2: Comparison between DUH, Gini, GSI, and SE



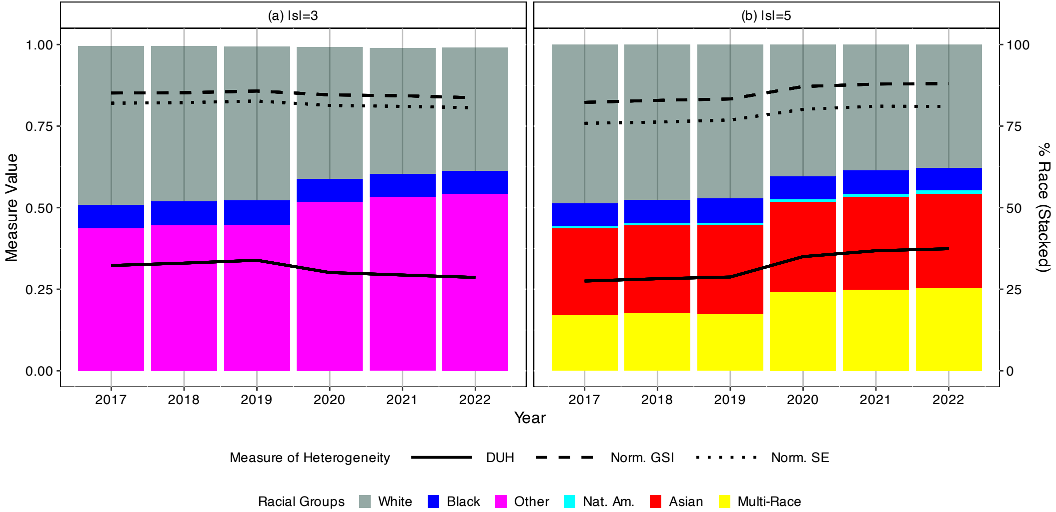
Reasonable Group Labels

While the satisfaction of GSYM implies that exact group labels are independent of the heterogeneity of a system, the satisfaction of CON implies that two systems are only directly comparable if the labeling of elements reflect a reasonable and interpretable grouping scheme.

Figure 3 illustrates this caveat with a practical example: the evolution of the racial composition of the San Francisco Metropolitan Statistical Area from 2017 to 2022, based on the American Community Survey (ACS) 1-year data (Ruggles et al., 2024).

Reasonable Group Labels

Figure 3: Different Group Labels Can Yield Opposite Inferences



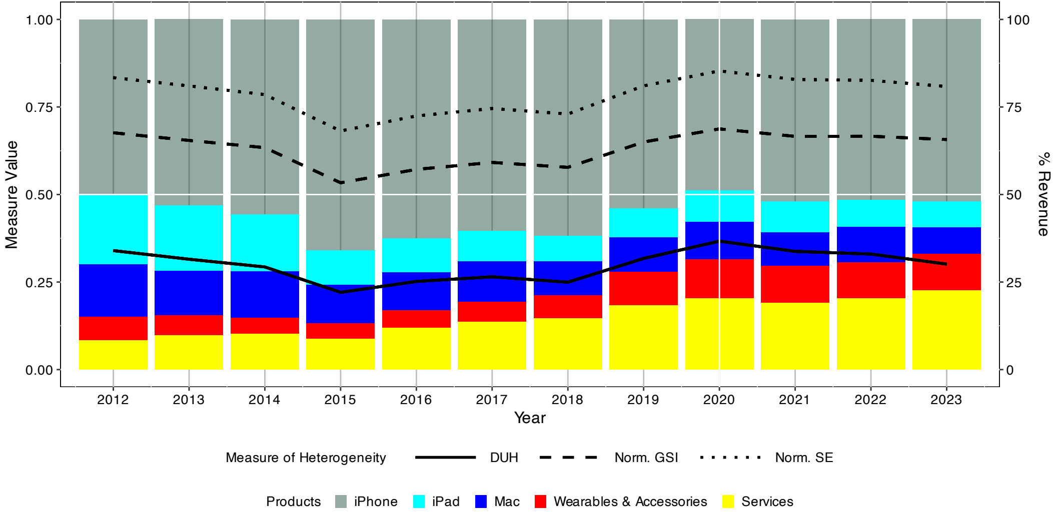
Empirical Example: Racial Heterogeneity

Table 1: The Progression of Racial Composition and Racial Heterogeneity of a Hypothetical City

Share (%)	Decade			
	1	2	3	4
White	60%	65%	66%	69%
Black	34%	26%	22%	17%
Other	6%	9%	11%	14%
DUH	0.235	0.257	0.284	0.315
Gini	0.460	0.440	0.450	0.450
GSI	0.521	0.502	0.499	0.475
Norm. GSI	0.781	0.753	0.749	0.713
Norm. SE	0.767	0.771	0.778	0.758

Empirical Example: Racial Heterogeneity

Figure 4: Revenue Heterogeneity using DUH, GSI, and SE



Thank You!

Bibliography I

Ruggles, Steven et al. (2024). *IPUMS USA: Version 15.0*. dataset. Minneapolis, MN. DOI: 10.18128/D010.V15.0. URL: https://doi.org/10.18128/D010.V15.0.

Characterization of DUH

Proof sketch

1. I show that any measure mapping an ordered system of group shares to the real numbers satisfies GSYM and INV.
2. I show that any Φ that satisfies GSYM, INV, and PPT must be a positive monotonic transformation of $\frac{1}{\hat{\sigma}_1}$.
3. I show that if $\Phi(\phi, \psi_p)$ satisfies INV, IND, and PPT, then ϕ and ψ_p must be multiplicatively separable. In other words, $\Phi = \phi \cdot \psi_p$.
4. I show that, holding $\hat{\sigma}_1$ constant and assuming GSYM, INV, and IND, the measure Φ, using the measure of evenness ψ_p as defined, satisfies PT if and only if $p > 1$ when $|s| > 3$ and $p \geq 1$ when $|s| = 3$.

Characterization of DUH

5. I show that if $\Phi(\phi, \psi_p)$ satisfies GSYM, INV, IND, and PPT, then $\phi = -c \cdot log_q\left(\hat{\sigma}_1\right), \, c, q \in \mathbb{R}_{++}$.
6. For the case of $p = 2$, using the Euclidean distance for ψ_p, I show that the DUH family satisfies PT, in addition to GSYM, INV, IND, and PPT, by taking the derivative of the extreme case in which $\hat{\sigma}_1$ is close to 1 and $\psi_2 = 1$ with respect to a transfer from the largest group to the second-largest group.
7. I show that the DUH family satisfies CON and UNI, in addition to GSYM, INV, IND, PT, and PPT.

[EXP] Expandability

Adding any arbitrary number of zero-groups does not affect the measure Φ.

$\Phi(n_1, \ldots, n_G)$ satisfies *Expandability* if

$$\Phi(n_1, \ldots, n_G) = \Phi(n_1, \ldots, n_G, 0)$$

[REP] Replication Principle

$\Phi(n_1, \ldots, n_G)$ satisfies *Replication Principle* (for concentration) if $\forall k \in \mathbb{N}$

$$\frac{1}{k}\Phi(n_1, \ldots, n_G) = \Phi\left(\underbrace{\frac{n_1}{k}, \frac{n_1}{k}, \ldots, \frac{n_1}{k}}_{\text{Sum to } n_1}, \frac{n_2}{k}, \frac{n_2}{k}, \ldots, \underbrace{\frac{n_G}{k}, \ldots, \frac{n_G}{k}}_{\text{Sum to } n_G}\right)$$

[SADD] Shannon's Additivity

Define $n_{gj} \geq 0$ such that $n_g = \sum_{j=1}^{m_g} n_{gj}$, $\forall g \in \{1, \ldots, G\}$, $\forall j \in \{1, \ldots, m_g\}$

$\Phi(n_1, \ldots, n_G)$ satisfies *Shannon's Additivity* if

$$\Phi(n_{11}, \ldots, n_{Gm_G}) = \Phi(n_1, \ldots, n_G) + \sum_{g=1}^{G} \frac{n_g}{n_S} \cdot \Phi\left(\frac{n_{g1}}{n_g}, \ldots, \frac{n_{gm_g}}{n_g}\right)$$

which implies (by setting $m_g = 1$, $\forall g \in \{1, \ldots, G-1\}$ and $n_{G'} = n_G + n_{G+1}$),

$$\begin{aligned}\Phi(n_1, \ldots, n_G, n_{G+1}) =& \Phi(n_1, \ldots, n_{G'}) \\ &+ \frac{n_{G'}}{n_1 + \ldots + n_{G-1} + n_{G'}} \cdot \Phi\left(\frac{n_G}{n_{G'}}, \frac{n_{G+1}}{n_{G'}}\right)\end{aligned}$$