A Friendly Amendment of the Theil Index

Duke University, March 2010

Kobi Abayomi [1] William Darity Jr. [2]

[1] Asst. Professor
Industrial Engineering - Statistics Group
Georgia Institute of Technology

[2] Department of Public Policy
Duke University and
Department of Economics
University of North Carolina

March 2010

Theil's Index

Theil's Index, a version of Shannon's Entropy, is introduced in econometrics as a measure for inequality. It is improperly specified for statistical use. We explore an adjustment of the Theil index by considering...

- Shannon's original Version
- Theil's (Mis)-Specification via Shannon
- A Friendly Amendment...
- A Test for Equi-Inequality?

Theil's Index

Theil's Index, a version of Shannon's Entropy, is introduced in econometrics as a measure for inequality. It is improperly specified for statistical use. We explore an adjustment of the Theil index by considering...

- Shannon's original Version
- Theil's (Mis)-Specification via Shannon
- A Friendly Amendment...
- A Test for Equi-Inequality?

Theil's Index

Theil's Index, a version of Shannon's Entropy, is introduced in econometrics as a measure for inequality. It is improperly specified for statistical use. We explore an adjustment of the Theil index by considering...

- Shannon's original Version
- Theil's (Mis)-Specification via Shannon
- A Friendly Amendment...
- A Test for Equi-Inequality?

Theil's Index

Theil's Index, a version of Shannon's Entropy, is introduced in econometrics as a measure for inequality. It is improperly specified for statistical use. We explore an adjustment of the Theil index by considering...

- Shannon's original Version
- Theil's (Mis)-Specification via Shannon
- A Friendly Amendment...
- A Test for Equi-Inequality?

Outline

Shannon's Axioms

Shannon represents an 'Information Source' - a random process on a discrete space - as a Markov process.
He supposes a "measure" H should have these qualities:

1. H should be continuous in event probability, p_i
2. For $p_i = p$, $\forall\ i$, H should monotonically increase
3. "If a choice be broken down, the original [H] should be the weighted sum"

Shannon's Axioms

Shannon represents an 'Information Source' - a random process on a discrete space - as a Markov process.
He supposes a "measure" H should have these qualities:

1. H should be continuous in event probability, p_i
2. For $p_i = p, \ \forall\ i$, H should monotonically increase
3. "If a choice be broken down, the original [H] should be the weighted sum"

Shannon's Axioms

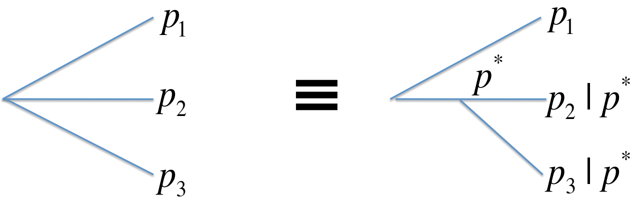
Shannon represents an 'Information Source' - a random process on a discrete space - as a Markov process.
He supposes a "measure" H should have these qualities:

1. H should be continuous in event probability, p_i
2. For $p_i = p, \ \forall\ i$, H should monotonically increase
3. "If a choice be broken down, the original [H] should be the weighted sum"

"If a choice be broken down"

Consistency over conditioning...

$$H(p_1, p_2, p_3) \equiv H(p_1, p^*) + p^* H(p_2|p^*, p_3|p^*)$$



Shannon's Theorem

The only H satisfying the three axioms is of the form (1949)

$$H = -K \sum_{i=1}^{n} p_i \, log \, p_i$$

Other 'Entropies'

- Gibbs Entropy: $S = -k_B \sum p_i \, log \, p_i$; (1872, 1878)
- von Neumann Entropy: $S - k_B Tr[\rho \, log_e \, \rho]$

Other 'Entropies'

- Gibbs Entropy: $S = -k_B \sum p_i \ log \ p_i$; (1872, 1878)
- von Neumann Entropy: $S - k_B Tr[\rho \ log_e \ \rho]$

Entropy's Career

Physics	Electrical Engineering → Computing → Computer Science
Statistics	Frechet → Ash → Kullback → Rissanen
Humanities	Theil: Econometrics

Entropy's Career

Physics	Electrical Engineering → Computing → Computer Science
Statistics	Frechet → Ash → Kullback → Rissanen
Humanities	Theil: Econometrics

Entropy's Career

Physics	Electrical Engineering → Computing → Computer Science
Statistics	Frechet → Ash → Kullback → Rissanen
Humanities	Theil: Econometrics

Outline

Theil's Version

A version of Theil's index is

$$T = n^{-1} \sum_{i=1}^{n} r_i \, log \, r_i \tag{1}$$

The probability of a particular event/realization is replaced with the income share for a particular element.

$$r_i = \frac{x_i}{\overline{x}} \tag{2}$$

Theil's Version

Theil's construction suggests

$$H = -\sum_{i=1}^{n} \frac{x_i}{n\bar{x}} \log_b \frac{x_i}{n\bar{x}} \tag{3}$$

This yields $\mathbf{p} = (p_1, ..., p_n) = (\frac{x_1}{n\bar{x}}, ..., \frac{x_n}{n\bar{x}}) = \mathbf{r}/n$ on a simplex, however, **the interpretation of Theil's index is inconsistent.**

Theil's Version

Since

$$\frac{log_{b_0}\ t}{log_{b_1}\ t} = K,\ \forall t \tag{4}$$

say, K can serve as a conversion between information units b_0 and b_1. **This feature is elided from Theil's construction.**

Theil's Decomposition

Thiel's index, on a population of n total individuals, is commonly defined as:

$$T = \sum_{j=1}^{m} p_j \; R_j \; log_b \; R_j + \sum_{j=1}^{m} p_j \; R_j \; T_j \tag{5}$$

with m disjoint groups, $g_1, ..., g_m$, each with n_j members - $n = \sum_j n_j$., and

$$T_j = n_j^{-1} \sum_{i \in g_j} r_{ij} \; log_b \; r_{ij}, \tag{6}$$

Decomposition

Consider this rewrite of (5),

$$T = \sum_G \frac{n_g}{n} \frac{\overline{X}_g}{\overline{\overline{X}}} log_b \frac{\overline{X}_g}{\overline{\overline{X}}} + \sum_G \frac{\overline{X}_g}{\overline{\overline{X}}} \frac{1}{n_g} \sum_g \frac{X_{ig}}{\overline{X}_g} log_b \frac{X_{ig}}{\overline{X}_g} \tag{7}$$

Decomposition

The first term on the right hand side

$$Across = \sum_G \frac{n_j}{n} \frac{\overline{X}_g}{\overline{X}} \, log_b \frac{\overline{X}_g}{\overline{X}} \tag{8}$$

is the measure of the *across* or *between* group inequality; the second term

$$Within = \sum_G \frac{\overline{X}_g}{\overline{X}} \frac{1}{n_g} \sum_g \frac{X_{ig}}{\overline{X}_g} log_b \frac{X_{ig}}{\overline{X}_g} \tag{9}$$

the *within* group inequality.

Misspecification

Examine the expectation of the *within* term:

$$E[Within] = E[\sum_G \frac{1}{\overline{X}n_g} \sum_g E[X_{ig} log_b \frac{X_{ig}}{\overline{X}_g} | G = g]] \tag{10}$$

$$= E[\sum_G \frac{1}{\overline{X}n_g} E[\sum_g X_{ig} log_b \frac{X_{ig}}{\overline{X}_g}]] \tag{11}$$

$$\geq E[\sum_G \frac{1}{\overline{X}n_g} E[n_g\overline{X}_g log_b \frac{n_g\overline{X}_g}{n_g\overline{X}_g}]] \tag{12}$$

$$\geq E[\sum_G \frac{1}{\overline{X}n_g} E[0]] = 0 \tag{13}$$

Misspecification

An inspection of the across term,

$$E[Across] = \sum_G \frac{n_g}{n}\frac{n_g}{n_g}E[\frac{X_{ig}}{\overline{X}}log_b\frac{\overline{X}_g}{\overline{X}}] \tag{14}$$

$$= \sum_G \frac{n_g}{n}E[\frac{X_{ig}}{\overline{X}}log_b\frac{\overline{X}_g}{\overline{X}}] \tag{15}$$

$$= \sum_G \frac{n_g}{n\mu}E[X_{ig}log_b\frac{\overline{X}_g}{\overline{X}}] \leq \sum_G \frac{n_g}{n\mu}E[X_{ig}] \cdot E[log_b\frac{\overline{X}_g}{\overline{X}}] \tag{16}$$

Misspecification

Notice

- $\frac{\overline{X}_g}{\overline{X}} < 1$ implies
- $log_b \frac{\overline{X}_g}{\overline{X}} < 0$, and $\uparrow$ as $b \uparrow$
- And $\frac{\overline{X}_g}{\overline{X}} > 1$ implies
- $log_b \frac{\overline{X}_g}{\overline{X}} > 0$, and $\downarrow$ as $b \uparrow$

Let $\epsilon_g^b \equiv E[log_b \frac{\overline{X}_g}{\overline{X}}]$

Misspecification

Notice

- $\frac{\overline{X}_g}{\overline{X}} < 1$ implies
- $log_b \frac{\overline{X}_g}{\overline{X}} < 0$, and $\uparrow$ as $b \uparrow$
- And $\frac{\overline{X}_g}{\overline{X}} > 1$ implies
- $log_b \frac{\overline{X}_g}{\overline{X}} > 0$, and $\downarrow$ as $b \uparrow$

Let $\epsilon_g^b \equiv E[log_b \frac{\overline{X}_g}{\overline{X}}]$

Misspecification

Notice

- $\frac{\overline{X}_g}{\overline{X}} < 1$ implies
- $log_b \frac{\overline{X}_g}{\overline{X}} < 0$, and $\uparrow$ as $b \uparrow$
- And $\frac{\overline{X}_g}{\overline{X}} > 1$ implies
- $log_b \frac{\overline{X}_g}{\overline{X}} > 0$, and $\downarrow$ as $b \uparrow$

Let $\epsilon_g^b \equiv E[log_b \frac{\overline{X}_g}{\overline{X}}]$

Misspecification

Notice

- $\frac{\overline{X}_g}{\overline{X}} < 1$ implies
- $log_b \frac{\overline{X}_g}{\overline{X}} < 0$, and $\uparrow$ as $b \uparrow$
- And $\frac{\overline{X}_g}{\overline{X}} > 1$ implies
- $log_b \frac{\overline{X}_g}{\overline{X}} > 0$, and $\downarrow$ as $b \uparrow$

Let $\epsilon_g^b \equiv E[log_b \frac{\overline{X}_g}{\overline{X}}]$

Misspecification

In these situations

$$\sum_G \frac{n_g}{n\mu} E[X_{ig}] \cdot E[log_b \frac{\overline{X}_g}{\overline{X}}]$$

$$\begin{aligned} &= \frac{1}{n\mu}\{n_1 E[X_1] \cdot (-\epsilon_1^b) && (17)\\ &+ \cdots n_{m-1} E[X_{m-1}] \cdot (-\epsilon_{m-1}^b) && (18)\\ &+ n_m E[X_m] \cdot (\epsilon_m^b)\} && (19)\\ &\leq 0 && (20) \end{aligned}$$

Misspecification

- Reminder to draw a picture, if I can
- This happens in data where few, small groups (n_g) have high, or low, income relative to population size.
- Effect is inflated, non-linearly, by choice of b
- While contribution, n_g, is only linear.

Misspecification

- Reminder to draw a picture, if I can
- This happens in data where few, small groups (n_g) have high, or low, income relative to population size.
- Effect is inflated, non-linearly, by choice of b
- While contribution, n_g, is only linear.

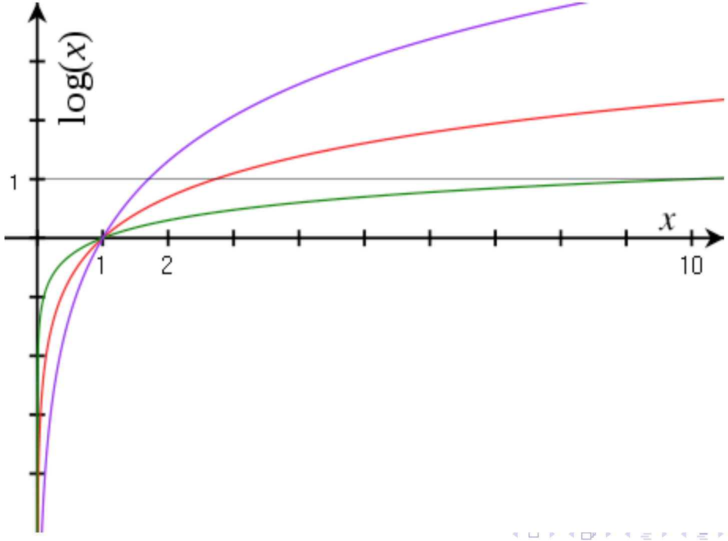
Misspecification

- Reminder to draw a picture, if I can
- This happens in data where few, small groups (n_g) have high, or low, income relative to population size.
- Effect is inflated, non-linearly, by choice of b
- While contribution, n_g, is only linear.

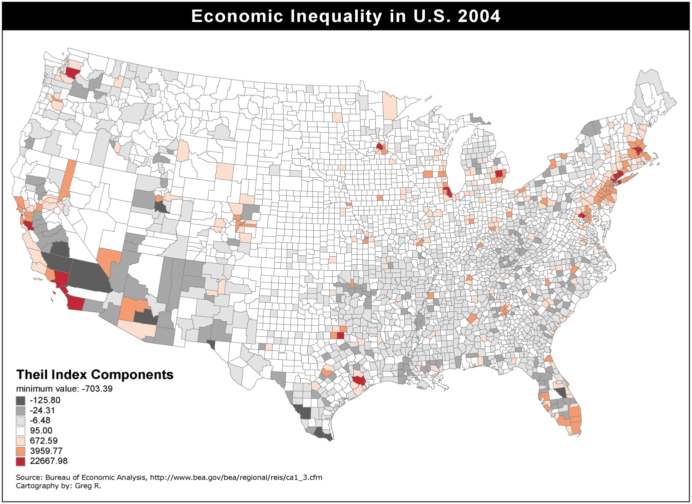
Misspecification

- Reminder to draw a picture, if I can
- This happens in data where few, small groups (n_g) have high, or low, income relative to population size.
- Effect is inflated, non-linearly, by choice of b
- While contribution, n_g, is only linear.

Illustration



Illustration



Comments

- The probability of a particular event/realization is replaced with the income share for a particular individual.
- We were measuring prob mass of events, now we are measuring the 'size' of individual
- In this sense: The individual is the event, and the income share is the prob mass

Comments

- The probability of a particular event/realization is replaced with the income share for a particular individual.
- We were measuring prob mass of events, now we are measuring the 'size' of individual
- In this sense: The individual is the event, and the income share is the prob mass

Comments

- The probability of a particular event/realization is replaced with the income share for a particular individual.
- We were measuring prob mass of events, now we are measuring the 'size' of individual
- In this sense: The individual is the event, and the income share is the prob mass

Comments

- The lack of a true event space (σ-algebra) yields a degenerate probability model...
- Though heuristic is consistent: *n*-dimensional simplex / Liouville family of distributions.
- But this belies a "deeper confusion" about Entropy and Probability [Jaynes (1965), *American Journal of Physics*].

Comments

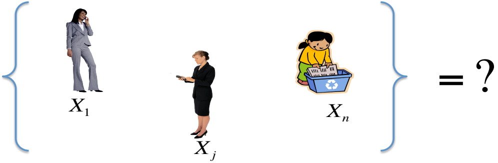
- The lack of a true event space (σ-algebra) yields a degenerate probability model...
- Though heuristic is consistent: n-dimensional simplex / Liouville family of distributions.
- But this belies a "deeper confusion" about Entropy and Probability [Jaynes (1965), *American Journal of Physics*].

Comments

- The lack of a true event space (σ-algebra) yields a degenerate probability model...
- Though heuristic is consistent: n-dimensional simplex / Liouville family of distributions.
- But this belies a "deeper confusion" about Entropy and Probability [Jaynes (1965), *American Journal of Physics*].

Illustration

$$\Omega = \left\{ E_1 \quad E_j \quad E_n \right\}$$



Shannon's Axiomatization

Notice:

$$T = log(n) - H$$

and that Shannon's original proof specified

$$H = -K \sum_{i=1}^{n} p_i \, log \, p_i$$

- Shannon's advice: K amounts to a choice of unit of measure (σ-algebra)
- Shannon chose the binary log (base $b = 2$)

Shannon's Axiomatization

Notice:

$$T = log(n) - H$$

and that Shannon's original proof specified

$$H = -K \sum_{i=1}^{n} p_i \, log \, p_i$$

- Shannon's advice: K amounts to a choice of unit of measure (σ-algebra)
- Shannon chose the binary log (base $b = 2$)

Skew-Sensitivity

One can arbitrarily choose partitions that guarantee across group sum dominates within groups sum.
Merely choose a collection G^* such that $x_j > n_j^{-1} \sum_{g_j} x_j$ for groups $g_1, ..., g_{[n/m]}$.
On this collection

$$A \equiv log \frac{n_j^{-1} \sum_{g_j} x_j}{n^{-1} \sum_G x_j} < log \frac{x_j}{n^{-1} \sum_G x_j} \equiv B$$

obscuring within group effects.

Log base characterizes 'Granularity'

As well, there exist collections G^* for every choice of b such that on G^* / g^*

$$A, B > 1$$

but

$$A, B < 1$$

on g^*

The Amnesia of (Probability) Measure

- The problems don't arise on Shannon's specification via the unit simplex
- Theil Index is on simplex of arbitrary sum
- Theil Index is an *ex parte* function on probability measures

The Amnesia of (Probability) Measure

- The problems don't arise on Shannon's specification via the unit simplex
- Theil Index is on simplex of arbitrary sum
- Theil Index is an *ex parte* function on probability measures

The Amnesia of (Probability) Measure

- The problems don't arise on Shannon's specification via the unit simplex
- Theil Index is on simplex of arbitrary sum
- Theil Index is an *ex parte* function on probability measures

Outline

Approaches

- **Adjust** b
- **Rank Transform** Estimate empirical CDF across and within groups and substitute for raw income shares

Approaches

- **Adjust** b
- **Rank Transform** Estimate empirical CDF across and within groups and substitute for raw income shares

Adjust b

In many settings — Rocke and Durbin Log-Linear hybridization; Box-Cox, Tukey families of transforms.. — b is a parameter to be estimated. $b = inf_g(R_g)$
Set

$$\hat{b} = min(r_g) = min_g(\frac{\overline{x}_g}{\overline{x}})$$

the minimum group income share.

Rank Transform

Essentially this is to revert Theil to Shannon...where $F_g = \mathbb{P}(group\ g)$ and $F_{X_g} = \mathbb{P}(X_g < x_g)$.
Set

$$Across \equiv \sum_g \hat{F}_g \cdot log_b(\hat{F}_g)$$

and

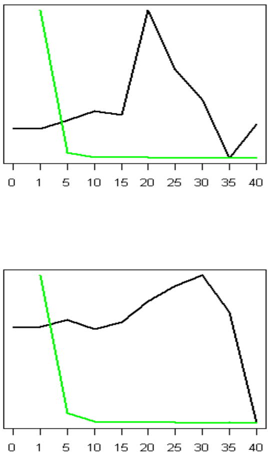
$$Within \equiv \sum_g \hat{F}_g \sum_{X_g} \hat{F}_{X_g} \cdot log(\hat{F}_{X_g})$$

Use $0 \leq \hat{F}_g, \hat{F}_{X_g} \leq 1$ or $\hat{F}_g, \hat{F}_{X_g} \in \{1, ..., m\}, \{1, ..., n_g\}$

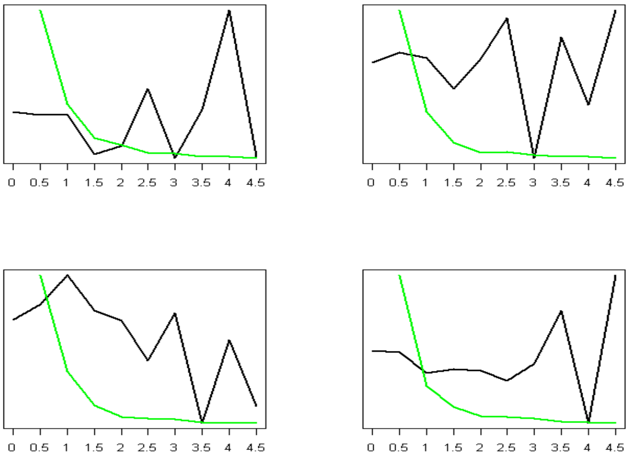
Some Plots

Plots of Thiel index (black) and Across/Within ratio for normal random variates mean zero. Standard deviations: 0 to 40 - figure (a), 0 to 4.5 - figure (b), 0 to 4.5 figure (c). The index is generally increasing, though not monotonically, due to random draws - figure (a). Figure (b) illustrates tradeoff between across and within terms at small departures from uniform within group distribution. In figure (c) the log base $b = min(R_j)$, the minimum income share. The plots illustrate varied across/within ratio behavior, due to randomness; there is support for a non-decreasing across/within ratio.

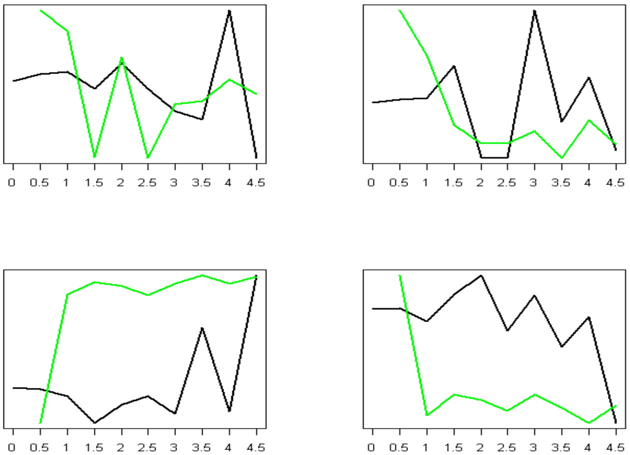
Groups of Equal Mean Uniform Discrete Random Variables



Groups of Equal Mean Uniform Discrete Random Variables



Groups of Equal Mean Uniform Discrete Random Variables, *b* is min.



Outline

Equi-Inequalities

Several hypothesis of Equi-Inequality to consider:

I Each group is equally unequal
II Groups are equal
III Each group contributes equally to inequality measure
IV Inequality Across Groups is equivalent to Inequality Within Groups

Equi-Inequalities

Several hypothesis of Equi-Inequality to consider:

I Each group is equally unequal

II Groups are equal

III Each group contributes equally to inequality measure

IV Inequality Across Groups is equivalent to Inequality Within Groups

Equi-Inequalities

Several hypothesis of Equi-Inequality to consider:

I Each group is equally unequal

II Groups are equal

III Each group contributes equally to inequality measure

IV Inequality Across Groups is equivalent to Inequality Within Groups

Equi-Inequalities

Several hypothesis of Equi-Inequality to consider:

I Each group is equally unequal

II Groups are equal

III Each group contributes equally to inequality measure

IV Inequality Across Groups is equivalent to Inequality Within Groups

Hypothesis I, say

Take hypothesis I, for example, i.e. $T_g = T \; \forall g$

$$\frac{Across}{Within} \rightarrow T^{-1}\left(\frac{\sum_g n_g \overline{X}_g log_b(\overline{X}_g)}{\sum_g n_g \overline{X}_g} - log_b(\overline{X})\right)$$

Hypothesis I, say

Since T, $log_b(\overline{X})$ and are ancillary under I, the test statistic of interest is

$$\Phi_I \equiv \frac{\sum_g n_g \overline{X}_g log_b(\overline{X}_g)}{\sum_g n_g \overline{X}_g}$$

Hypothesis II, say

This is equivalent to the ordinary ANOVA or Bartlett's test for means or variances.

Hypothesis III, say

This is a generalization of I; the test statistic of interest is

$$\Phi_{III} = \sum_g \frac{n_g}{n} \frac{\overline{X}_g}{\overline{\overline{X}}} log_b(\frac{\overline{X}_g}{\overline{\overline{X}}})$$

Under III

$$\Phi_{III} \rightarrow m \cdot C$$

Hypothesis IV

This is the most general hypothesis; the test statistic is the ratio of across to within

$$\Phi_{IV} = \frac{\sum_G \frac{n_g}{n} \frac{\overline{X}_g}{\overline{\overline{X}}} \; log_b \frac{\overline{X}_g}{\overline{\overline{X}}}}{\sum_G \frac{\overline{X}_g}{\overline{\overline{X}}} \frac{1}{n_g} \sum_g \frac{X_{ig}}{\overline{X}_g} log_b \frac{X_{ig}}{\overline{X}_g}}$$

If b is well chosen, this a ratio of two positive quantities that is 1 at the null hypothesis. Similar to an F statistic.

Significance, Next

- Places Theil's contribution in statistical context, distributions on simplex (Dirichlet, Liouville)
- Codifies the index as a measure of distributional dependence
- Statistical properties/Hypothesis Tests are necessary descriptives for inequality, against inequality. 'Significance'
- Apply to HRS data: tests of change in inequality over time.

Significance, Next

- Places Theil's contribution in statistical context, distributions on simplex (Dirichlet, Liouville)
- Codifies the index as a measure of distributional dependence
- Statistical properties/Hypothesis Tests are necessary descriptives for inequality, against inequality. 'Significance'
- Apply to HRS data: tests of change in inequality over time.

Significance, Next

- Places Theil's contribution in statistical context, distributions on simplex (Dirichlet, Liouville)
- Codifies the index as a measure of distributional dependence
- Statistical properties/Hypothesis Tests are necessary descriptives for inequality, against inequality. 'Significance'
- Apply to HRS data: tests of change in inequality over time.

Significance, Next

- Places Theil's contribution in statistical context, distributions on simplex (Dirichlet, Liouville)
- Codifies the index as a measure of distributional dependence
- Statistical properties/Hypothesis Tests are necessary descriptives for inequality, against inequality. 'Significance'
- Apply to HRS data: tests of change in inequality over time.