Let's get this programming task nailed down

JP Ferguson

March 3, 2019

1 Hierarchical decomposition of Theil's information statistic

1.1 The two-level case

If our unit has G different groups, each with J_g sub-units, then we have two sources of segregation: the uneven distribution of races between the G groups, and the uneven distribution of races between the J_g sub-units within each group g. We want to account for these separately. We calculate H_g for between-group segregation. We then calculate within-group, between-sub-unit segregation for each g, which we denote H_j^g. Total segregation, which we denote $H_{\langle GJ \rangle}$, is between-group segregation plus the weighted sum of within-group segregations:

$$
\begin{aligned}
H_{\langle GJ \rangle} &= \overbrace{\sum_g \frac{p_g}{p} \frac{E - E_g}{E}}^{\text{Between group}} + \underbrace{\sum_g \frac{p_g}{p} \frac{E_g}{E} \left(\sum_{j \in g} \frac{p_{gj}}{p_g} \frac{E_g - E_{gj}}{E_g} \right)}_{\text{Within group, between sub-unit}} \\
&= H_g + \sum_g w_g d_g h_g \\
&= H_g + \bar{H}_j^g
\end{aligned} \tag{1}
$$

Note first that we denote any one group's contribution to total segregation as H_j^g and the sum across groups' contributions as $\bar{H}_j^g$. For convenience, we denote the relative size, or weight, of a group w_g; the relative entropy, or diversity, of a group d_g; and the segregation between sub-units within the group h_g. (**?** offer a proof of this identity.) Note second that, since this a group decomposition, we have the second weighting term d_g. In any hierarchical decomposition you have to include this weight at all levels above the lowest. Note finally that h_g is a recursive definition of H; thus this process could be repeated for sub-groups in G.

1.2 The three-level case

For clarity, rather than talk about groups and sub-groups, we will talk about areas and groups, but note that groups nest within areas and that establishments nest within groups. This gives us three sources of segregation: the uneven distribution of races between the A areas, the uneven distribution of races between the G_a groups within each area, and the uneven distribution of races between the J_{ag} sub-units within each area-group. Building on the logic above, we calculate H_a for between-area segregation. We then calculate within-area, between-group segregation for each a, which we denote H_g^a. Finally we calculate within-area-group, between-establishment segreagtion for each g_a, which we denote H_j^{ag}. Total segregation, which we deonte $H_{\langle AGJ \rangle}$, is between-area segregation plus the ewighted sum of within-area segregations, plus the weighted sum of within area-group segregations:

$$
\begin{aligned}
H_{\langle AGJ \rangle} &= \overbrace{\sum_a \frac{p_a}{p} \frac{E - E_a}{E}}^{\text{Between area}} + \overbrace{\sum_a \frac{p_a}{p} \frac{E_a}{E} \left(\sum_g \frac{p_{ag}}{p_a} \frac{E_{ag}}{E_a} \frac{E_a - E_{ag}}{E_a} \right)}^{\text{Within area}} \\
&\quad + \overbrace{\sum_a \frac{p_a}{p} \frac{E_a}{E} \left(\sum_g \frac{p_{ag}}{p_a} \frac{E_{ag}}{E_a} \left(\sum_j \frac{p_{agj}}{p_{ag}} \frac{E_{agj}}{E_{ag}} \frac{E_{ag} - E_{agj}}{E_{ag}} \right) \right)}^{\text{Within area-group}} \\
&= H_a + \sum_a w_a d_a h_a + \sum_a w_a d_a \left(\sum_g w_{ag} d_{ag} \left(\sum_j h_{agj} \right) \right) \\
&= H_a + \bar{H}_g^a + \bar{H}_j^{ag}
\end{aligned}
\tag{2}
$$

In the within-area-group segregation term there are three summations. The first distributes over the second one, such that a single group's contribution, H_j^{ag}, is $\sum_a w_a d_a w_{ag} d_{ag} \left(\sum_j h_{agj} \right)$. In other words, the contribution of a group that is nested within areas is the relative size- and diversity-weighted sum of the group's contribution in each area, weighted by the relative size and diversity of each area.

2 Mechanism decomposition of the statistic

2.1 The two-level case

We are interested in whether changes within and between groups, such as industries, can help explain changes in employment segregation over time. Within the context of the Theil statistic, this maps to changes in the group components of $\bar{H}_j^g$ in the two-level case. We have shown that, for any single group $g \in G$, $H_j^g = w_g d_g h_g$. Because each group's contribution to $\bar{H}_j^g$ is a product of three terms, each of which can vary independently of the others, it can be useful to separate changes in H_j^g into changes in each of those three terms. We refer to this decomposition of the *dynamics* of H as a "mechanism" decomposition to distinguish it from the hierarchical, cross-sectional decomposition detailed above.

We prefer to approach this with some geometric intuition. A product of three terms can be visualized as a rectangular solid in $\mathcal{R}^3$ whose dimensions are w_g, d_g, and h_g; see Figure 1. Adding a subscript to track time (and assuming we understand $g_t \in G_t$) yields $H_j^{g_t} = w_{g_t} d_{g_t} h_{g_t}$. Change in these components between two time periods is then the difference in two products:

$$
\forall g \in G : \Delta H_j^g = w_{g_2} d_{g_2} h_{g_2} - w_{g_1} d_{g_1} h_{g_1} \tag{3}
$$

Because $x_{g_{t+1}} = x_{g_t} + \Delta x$ for $x \in \{w, d, h\}$, we can analyze ΔH_j^g into the difference in volume between two rectangular solids, as in Figure 2.[1]

The geometry of Figure 2 clarifies why change in a group's contribution to total segregation can and should have *seven* components. Three of these are like main effects. For example, Δwdh

[1] In Figure 2 we draw all changes as positive, but the math works the same if one or more are negative.

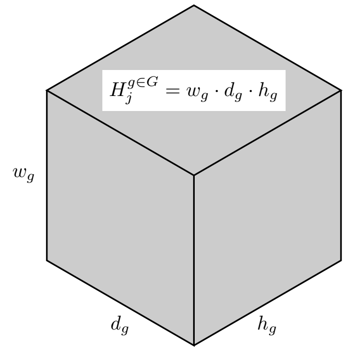

Figure 1: The components of a group's contribution to overall segregation H_j^g, visualized as a solid in $\mathcal{R}^3$

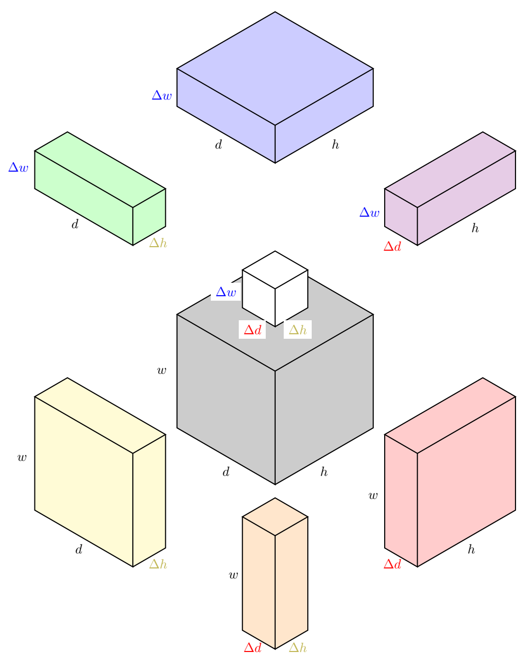

Figure 2: Analyzing ΔH_j^g into the difference of two rectangular solids. We omit g sub-scripts for clarity. The eight components form a solid whose dimensions are $w+\Delta w$, $d+\Delta d$, and $h+\Delta h$, which corresponds to $H^g_{j_{t+1}}$. Subtracting wdh, which corresponds to $H^g_{j_t}$, gives the between-period change, which can then be analyzed as shown.

represents the effect of increasing the relative size of a group while leaving diversity and segregation within the group unchanged. Similarly, $w\Delta dh$ is the effect of increasing diversity within g while keeping its size and segregation between its establishments unchanged. Another three components resemble two-way interactions: thus, $\Delta w\Delta dh$ captures how changing size and changing diversity within g tend to go hand-in-hand. Finally, $\Delta w\Delta d\Delta h$ is the "messy residual," the portion of increasing segregation that, due to simultaneous movement, cannot be assigned between these three channels. We track these subscripts to identify these components. Let $\underset{x}{\Delta} H_j^g$ be the change in H_j^g from time t to time $t+1$ that is due to change in component(s) x. Thus for example $\underset{h}{\Delta} H_j^g$ represents $wd\Delta h$, $\underset{dh}{\Delta} H_j^g$ represents $w\Delta d\Delta h$ and so on.

This analysis of changes in group contributions to segregation into size, diversity, and segregation components parallels a similar discussion of the "causes" of changing segregation in **?**. We have adapted our formal approach here from economic work on changes in aggregate firm producivity; see for example **?**. We detail those connections more explicitly, and include the full proof that these components sum to changes in total segregation, in Appendix **??**.

2.2 The three-level case

As with the two-level case, let $\underset{x}{\Delta} H_j^{ag}$ be the change in H_j^{ag} from time t to time $t+1$ that is due to change in component(s) x. Consider the simplest case where only one component, like h_{ag}, changes, and note that $h_{ag_{t+1}} \equiv h_{ag_t} + \Delta h_{ag}$. Then we get the following:

$$
\begin{aligned}
\underset{h}{\Delta} H_j^{ag} &= H_{j_{t+1}}^{ag} - H_{j_t}^{ag} \\
&= \sum_a w_{a_t} d_{a_t} w_{ag_t} d_{ag_t} \left(h_{ag_t} + \Delta h_{ag}\right) - \sum_a w_{a_t} d_{a_t} w_{ag_t} d_{ag_t} h_{ag_t} \\
&= \sum_a w_{a_t} d_{a_t} w_{ag_t} d_{ag_t} \Delta h_{ag} \\
&\quad + \sum_a \left(w_{a_t} d_{a_t} w_{ag_t} d_{ag_t} h_{ag_t} - w_{a_t} d_{a_t} w_{ag_t} d_{ag_t} h_{ag_t}\right) \\
&= \sum_a w_{a_t} d_{a_t} w_{ag_t} d_{ag_t} \Delta h_{ag}
\end{aligned} \tag{4}
$$

This holds for any of the components in H_j^{ag}. In other words, for the three-level case, to derive the contributions of groups that are nested in areas, we can add up group contributions within area-groups, then do a size- and diversity-weighted sum across areas. This implies that mechanism decomposition follows the same recursively weighted summation that hierarchical decomposition does.

3 Calculation of mechanism decomposition

3.1 Design considerations

Defining standard errors for any mechanism changes in group contributions to the Theil statistic is not done analytically, because there is no agreed procedure for doing so (**?**). Instead we simulate probability distributions through bootstrap and permutation tests. These simulations imply calculating each relevant statistic many thousands of times, which in turn impels fast calculation. We found that we could speed up calculation considerably by implementing the statistics as a series of transformations to a design matrix. Our goal is to generate probability distributions for year-on-year changes, but for any year t we only need that year's data to calculate statistics.

Assume N sub-units that are members of a set of G mutually-exclusive and completely exhaustive groups. These groups are themselves present within A areas. For any year, we define the design matrix $\mathbf{Q}$ that has the area a, group g, sub-unit n, and a set of vectors $\mathbf{R}$ with counts of r different races:

$$\begin{bmatrix} 011 & 1 & aaa & 10 & 9 & 30 & 2 \\ 011 & 1 & aab & 4 & 4 & 4 & 4 \\ 011 & 2 & bbb & 30 & 19 & 0 & 3 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 999 & 6 & zzz & 12 & 17 & 22 & 19 \end{bmatrix} \tag{5}$$

We also specify several operators (matrices) that we will use repeatedly to calculate statistics:

- **The j-operator**: For any $n \times k$ matrix $\mathbf{A}$, $\mathbf{Aj}$ is defined as post-multiplication by a $k \times 1$ vector of 1s. Thus $\mathbf{Aj}$ produces a column vector with the row totals of $\mathbf{A}$.

- **The i-operator**: For any $n \times k$ matrix $\mathbf{A}$, $\mathbf{iA}$ is defined as pre-multiplication by a $1 \times n$ vector of 1s. Thus $\mathbf{iA}$ produces a row vector with the column totals of $\mathbf{A}$.

- **The S-operator**: For any $n \times k$ matrix $\mathbf{A}$ with a vector $\mathbf{s}$ that defines groups $g \in G$ of observations, $\mathbf{s_g}$ is a $n \times 1$ vector whose ith element equals 1 if $a_{i,s} = g$ and equals 0 otherwise. In turn, $\mathbf{S_g}$ is a $n \times G$ matrix, where G is the number of groups. The vector $\mathbf{s}$ can be a unique vector in $\mathbf{A}$ or it can be a vector with unique values for each tuple of values across multiple vectors.

- **The Hadamard product**: For any matrices $\mathbf{A}$ and $\mathbf{B}$ of equal size, the Hadamard product $\mathbf{A} \odot \mathbf{B}$ is a binary operator whose (i, j)th element is $a_{i,j} \cdot b_{i,j}$. Thus the Hadamard operator $\odot$ performs element-wise matrix multiplication.

- **The Hadamard quotient**: For any matrices $\mathbf{A}$ and $\mathbf{B}$ of equal size, the Hadamard quotient $\mathbf{A} \oslash \mathbf{B}$ is a binary operator whose (i, j)th element is $a_{i,j}/b_{i,j}$. Thus the Hadamard operator $\oslash$ performs element-wise matrix division.

- **The $\triangle$-operator**: For any $n \times k$ matrix $\mathbf{A}$, $\mathbf{A}^{\triangle}$ is an $n \times k$ matrix whose (i, j)th element is $\ln(a_{i,j})$. For any element of $\mathbf{A}$ that equals 0, the corresponding element of $\mathbf{A}^{\triangle}$ is defined as 0.

3.2 The two-level case

3.2.1 Generating w_g and w_{gj}

Recall that $w_g = p_g/p$ and $w_{gj} = p_{gj}/p_g$. Because $\mathbf{R}$ has each sub-unit's racial counts as row elements, $\mathbf{Rj}$ is the vector holding p_{gj}, the total size of each sub-unit.

To calculate p_g, that is, to generate total size for the higher-level groups rather than for the sub-units, apply the S-operator. For example, $\mathbf{S_g}(\mathbf{S_g}'\mathbf{Rj})$ yields an $N \times 1$ vector with each sub-unit's relevant group total. We can then calculate w_{gj} with the Hadamard quotient:

$$w_{gj} = \overset{N\times r}{\mathbf{R}} \overset{r\times 1}{\mathbf{j}} \oslash \overset{N\times G}{\mathbf{S_g}} (\overset{G\times N}{\mathbf{S_g}'} \overset{N\times r}{\mathbf{R}} \overset{r\times 1}{\mathbf{j}}) \tag{6}$$

The parentheses indicating order of operations are important here. $\mathbf{SS'}$ produces an $N \times N$ matrix. When, as here, $N \gg 100,000$, this matrix can easily comprise tens of billions of elements. Thus it is important to post-multiply $\mathbf{S'}$ by successive terms before pre-multiplying it by $\mathbf{S}$.

Calculating p works similarly, except that we do not need to segment the summation by groups. Instead we can use the i-operator to sum across all observations, and thus calculate the denominator of w_g. The numerator meanwhile is the denominator from w_{gj}:

$$w_g = \overset{N\times G}{\mathbf{S_g}} \; (\overset{G\times N}{\mathbf{S_g}'} \overset{N\times r}{\mathbf{R}} \overset{r\times 1}{\mathbf{j}}) \oslash \overset{N\times 1}{\mathbf{i}'} \; (\overset{1\times N}{\mathbf{i}} \overset{N\times r}{\mathbf{R}} \overset{r\times 1}{\mathbf{j}}) \tag{7}$$

3.2.2 Generating E

For any sub-unit j, $E_{gj} = -\sum_r \phi_r \ln \phi_r$, where ϕ_r is race r's share of j's total population. We can generate $\mathbf{\Phi_{gj}}$, which expresses $\mathbf{R}$ as shares of a total rather than counts:

$$\mathbf{\Phi_{gj}} = \overset{N\times r}{\mathbf{R}} \oslash \overset{N\times r}{\mathbf{R}} \overset{r\times 1}{\mathbf{j}} \overset{1\times r}{\mathbf{j}'} \tag{8}$$

Then we can calculate a vector of E_{gj}s as $-1 \times (\mathbf{\Phi_{gj}} \odot \mathbf{\Phi_{gj}^{\triangle}})\mathbf{j}$.

To calculate E_g we use the S-operator again. If $\mathbf{\Phi_g} = \mathbf{S_g S_g' R} \oslash \mathbf{S_g S_g' R j j'}$, then $E_g = -1 \times (\mathbf{\Phi_g} \odot \mathbf{\Phi_g^{\triangle}})\mathbf{j}$.

3.2.3 Generating h_g

The h terms are undefined at the sub-unit level, so there is only the internal segregation within groups to calculate. This is just the segregation between establishments in a group, and all of the components for this calculation have already been defined:

$$h_g = \mathbf{S_g}(\mathbf{S_g}'(w_{gj} \odot ((E_g - E_{gj}) \oslash E_g))) \tag{9}$$

3.3 The three-level case

3.3.1 Generating w_a, w_{ag} and w_{agj}

The populations and weights for a three-level Theil decomposition are calculated in the same was as in the two-level case, with one additional layer:

$$\begin{aligned} p_{agj} &= \mathbf{Rj} \\ p_{ag} &= \mathbf{S_{ag}}(\mathbf{S'_{ag}}\mathbf{Rj}) \\ p_a &= \mathbf{S_a}(\mathbf{S'_a}\mathbf{Rj}) \\ p &= \mathbf{i}'(\mathbf{iRj}) \end{aligned} \tag{10}$$

$$\begin{aligned} w_{agj} &= p_{agj} \oslash p_{ag} \\ w_{ag} &= p_{ag} \oslash p_a \\ w_a &= p_a \oslash p \end{aligned} \tag{11}$$

3.3.2 Generating E and d

Measures of entropy at different levels also works similarly to the two-level case:

$$\begin{aligned}\mathbf{\Phi_{agj}} &= \mathbf{R} \oslash \mathbf{Rjj'} \\ \mathbf{\Phi_{ag}} &= \mathbf{S_{ag}}(\mathbf{S'_{ag}R} \oslash (\mathbf{S_{ag}}(\mathbf{S'_{ag}Rjj'}))) \\ \mathbf{\Phi_{a}} &= \mathbf{S_{a}}(\mathbf{S'_{a}R} \oslash (\mathbf{S_{a}}(\mathbf{S'_{a}Rjj'}))) \\ \mathbf{\Phi} &= \mathbf{i'}(\mathbf{iR} \oslash (\mathbf{i'}(\mathbf{iRjj'})))\end{aligned} \tag{12}$$

Likewise,

$$\begin{aligned}E_{agj} &= -1 \times (\mathbf{\Phi_{agj}} \odot \mathbf{\Phi^{\triangle}_{agj}})\mathbf{j} \\ E_{ag} &= -1 \times (\mathbf{\Phi_{ag}} \odot \mathbf{\Phi^{\triangle}_{ag}})\mathbf{j} \\ E_{a} &= -1 \times (\mathbf{\Phi_{a}} \odot \mathbf{\Phi^{\triangle}_{a}})\mathbf{j} \\ E &= -1 \times (\mathbf{\Phi} \odot \mathbf{\Phi^{\triangle}})\mathbf{j}\end{aligned} \tag{13}$$

In cases of three or more levels, all aggregations above the lowest require a weight for relative entropy:

$$\begin{aligned}d_{agj} &= E_{agj} \oslash E_{ag} (\text{Not used}) \\ d_{ag} &= E_{ag} \oslash E_{a} \\ d_{a} &= E_{a} \oslash E\end{aligned} \tag{14}$$

Generating h_{ag}

As in the two-level case, h is not defined at the lowest level of the decomposition. We also do not have to generate h_a. In a hierarchical structure where establishments are nested into area-groups that are nested into areas, h_a gives the between-area component of segreagtion. To get individual group contributions, Instead of moving from h_{ag} to h_a, we should move from h_{ag} to h_g. We will do this by summing area-groups *by* group *across* areas, weighting each component of the sum by its areas relative size and diversity. We nonetheless show the calculation of h_a here for completeness' sake.

$$\begin{aligned}h_{agj} &= w_{agj} \odot ((E_{ag} - E_{agj}) \oslash E_{ag}) \\ h_{ag} &= \mathbf{S_{ag}}(\mathbf{S'_{ag}} h_{agj}) \\ h_{a} &= \mathbf{S_{a}}(\mathbf{S'_{a}}(w_{ag} d_{ag} h_{ag})) \\ h_{g} &= \mathbf{S_{g}}(\mathbf{S'_{g}}(w_{a} d_{a} w_{ag} d_{ag} h_{agj}))\end{aligned} \tag{15}$$

Permutations

If we get all of this right, then permutations are relatively simple. All we have to do is separate the data by labor markets, then within labor markets shuffle establishments across groups (economic sectors) in a way that preserves total employment as much as possible within groups. Then we do the same calculations, eventually deriving counterfactual values for $\underset{h}{\Delta} H^{ag}_{j}$ and $\underset{e}{\Delta} H^{ag}_{j}$. We repeat that however many times, and we have our permutation test.