Measuring Segregation: Proofs of Miscellaneous Claims

David Frankel and Oscar Volij

March 17, 2007

The following claim will be used to verify that an index satisfies Continuity.

Claim 1 *Any index that is a continuous function of the T_g^n's (the numbers of each group g in each school n) represents an ordering that satisfies the axiom of Continuity.*

Proof. Let S be a segregation index. Assume that S is a continuous function of the T_g^n's. Let X, Y, and Z be three districts. Define the function $f : [0,1] \to \mathbb{R}$ by

$$f(c) = S(cX \uplus (1-c)Y)$$

Since S is continuous in the T_g^n, f is continuous in c. The sets $(-\infty, S(Z)]$, and $[S(Z), \infty)$ are closed in $\mathbb{R}$. Consequently, $f^{-1}((-\infty, S(Z)])$ and $f^{-1}([S(Z), \infty))$ are closed in $[0,1]$ (For continuous functions, their inverse image of closed sets are closed). But these are just the sets $\{c \in [0,1] : S(Z) \geq S(cX \uplus (1-c)Y)\}$ and $\{c \in [0,1] : S(cX \uplus (1-c)Y) \geq S(Z)\}$, respectively. Q.E.D.

Claim 2 *The multigroup dissimilarity index satisfies all properties but IND1, IND2, GDP, SSD, and SGD. It satisfies SI only in the two-group case.*

Proof. By inspection, D satisfies WSI, N, and SYM. It satisfies C by Claim 1. For SDP, let $X \in C$ be a district and let n be a school in X. Let X' be the district that results from dividing n into two schools, n_1 and n_2, and let $\alpha = T^{n_1}/T^n$. Then,

$$\begin{aligned} D(X') - D(X) &= \frac{1}{2I} \sum_{g \in \mathbf{G}(X)} \left(P^{n_1} \left| p_g^{n_1} - P_g \right| + P^{n_2} \left| p_g^{n_2} - P_g \right| - P^n \left| p_g^n - P_g \right| \right) \\ &= \frac{1}{2I} \sum_{g \in \mathbf{G}(X)} P^n \left(\alpha \left| p_g^{n_1} - P_g \right| + (1-\alpha) \left| p_g^{n_2} - P_g \right| - \left| p_g^n - P_g \right| \right) \end{aligned}$$

Since $p_g^n = \alpha p_g^{n_1} + (1-\alpha) p_g^{n_2}$ and absolute value is a convex function, $D(X') - D(X) \geq 0$. Moreover, if n_1 and n_2 have the same group distributions, $D(X') - D(X) = 0$. To see that D violates IND1, consider the following districts: $X = \{(9,5),(1,5)\}$, $Y = \{(5,9),(5,1)\}$ and $Z = \{(6,4)\}$. Districts X and Y have the same population and group distribution. By symmetry of D, $D(X) = D(Y)$. However, it can be checked that $D(X \uplus Z) \neq D(Y \uplus Z)$. As for IND2, consider the districts $X = \langle (2,8) \rangle$, $Y = \langle (8,2) \rangle$ and $Z = \langle (1,5),(9,5) \rangle$. It can be checked that $D(X \uplus Z) \neq D(Y \uplus Z)$ even though $D(X \uplus c(Z)) = D(Y \uplus c(Z))$, so D violates IND2 as well. As for GDP, consider the district $X = \langle (0,10),(10,0) \rangle$ and the district $Y = \langle (0,0,5,5),(5,5,0,0) \rangle$ that is obtained after we split evenly each of X's groups. It can be checked that $D(X) \neq D(Y)$. Q.E.D.

Claim 3 *The multigroup Gini index satisfies all properties but IND1, IND2, GDP, SSD, and SGD. It satisfies SI only in the two-group case.*

Proof. By inspection, G satisfies WSI, N, and SYM. It satisfies axiom C by Claim 1. We now check whether G satisfies SDP. Let $X \in C$ be a district and let n be a school in X. Let X' be the district that results from dividing n into two schools, n_1 and n_2. We must show that $G(X') \geq G(X)$, with equality if the two schools have the same group distribution. But

$$\begin{aligned} G(X') - G(X) &= \frac{1}{I} \sum_{g=1}^{G} \frac{T^{n_1} T^{n_2}}{TT} \left| \frac{T_g^{n_1}}{T^{n_1}} - \frac{T_g^{n_2}}{T^{n_2}} \right| \\ &+ \frac{1}{I} \sum_{g=1}^{G} \sum_{\substack{m=1,\ldots,N \\ m \neq n}} \left(\begin{array}{c} \frac{T^m T^{n_1}}{TT} \left| \frac{T_g^m}{T^m} - \frac{T_g^{n_1}}{T^{n_1}} \right| + \frac{T^m T^{n_2}}{TT} \left| \frac{T_g^m}{T^m} - \frac{T_g^{n_2}}{T^{n_2}} \right| \\ - \frac{T^m T^n}{TT} \left| \frac{T_g^m}{T^m} - \frac{T_g^n}{T^n} \right| \end{array} \right) \end{aligned}$$

The first sum is nonnegative. The summand in the second line equals

$$\frac{T^m}{TT} \left(\left| T^{n_1} \frac{T_g^m}{T^m} - T_g^{n_1} \right| + \left| T^{n_2} \frac{T_g^m}{T^m} - T_g^{n_2} \right| - \left| T^n \frac{T_g^m}{T^m} - T_g^n \right| \right)$$

The arguments of the first two absolute value functions sum to the argument of the third absolute value function. However, absolute value is a convex function. Hence, the summand is nonnegative for all g. Moreover, if the two schools have the same group distributions, then the arguments of the three absolute value functions are proportional to each other and thus all of the same sign. So the summand is zero. This shows that G satisfies SDP. The same counterexamples used in the proof of Claim 2 show that G violates IND1, IND2, and GDP. Q.E.D.

Claim 4 *The Entropy index satisfies all properties but IND2, GDP, SI, SSD, and SGD.*

Proof. Given the relation between H and the Mutual Information index, it is clear that H satisfies WSI, SDP, N, SYM, and C. As for IND1, note that if X, Y are two districts with equal populations and equal group distributions, then (a) $E(X) = E(Y)$ and (b) $X \uplus Z$ and $Y \uplus Z$ have equal group distributions for all Z. Fact (b) implies that $E(X \uplus Z) = E(Y \uplus Z)$. Accordingly, $H(X) - H(Y) = \frac{M(X) - M(Y)}{E(Y)}$ and $H(X \uplus Z) - H(Y \uplus Z) = \frac{M(X \uplus Z) - M(Y \uplus Z)}{E(X \uplus Y)}$. Hence, H satisfies IND1 as M does. As for IND2, consider the districts $X = \langle (2,0,0), (0,2,0), (0,0,2) \rangle$, $Y = \langle (3,0,0), (0,3,0) \rangle$ and $Z = \langle (1,0,0), (0,1,0) \rangle$. It can be checked that $H(X \uplus Z) = H(Y \uplus Z)$ even though $H(X \uplus c(Z)) \neq H(Y \uplus c(Z))$, so H violates IND2 as well. As for GDP, consider the district $X = \langle (0,10), (10,0) \rangle$ and the district $Y = \langle (0,0,5,5), (5,5,0,0) \rangle$ that is obtained after we split evenly each of X's groups. It can be checked that $H(X) \neq H(Y)$. Q.E.D.

Claim 5 *The Normalized Exposure index satisfies all properties but GDP, SI, IND2, SSD, and SGD. It satisfies IND1 only in the two-group case.*

Proof. By inspection, Normalized Exposure satisfies WSI and SYM. It satisfies axiom C by Claim 1. To see that it satisfies SDP, suppose X' results from splitting school n in X into two schools, n_1 and n_2. Then

$$P(X') - P(X) = \sum_{g=1}^{G} \frac{1}{1-P_g} \left[P^{n_1}(p_g^{n_1} - P_g)^2 + P^{n_2}(p_g^{n_2} - P_g)^2 - P^n(p_g^n - P_g)^2 \right]$$

But $P^{n_1}(p_g^{n_1} - P_g) + P^{n_2}(p_g^{n_2} - P_g) = P^n(p_g^n - P_g)$ and square is a convex function so the weak inequality in SDP holds. Moreover, if schools n_1 and n_2 have the same group distribution, then $p_g^{n_1} = p_g^{n_2} = p_g^n$ so the sum is identically zero. So SDP holds. To see why IND1 is satisfied with two groups, let $X, Y \in C$ be two districts with two groups, equal total size, and equal group distributions. Then for all districts $Z \in C$ with the same set of groups, we must show that $P(X) \geq P(Y)$ if and only if $P(X \uplus Z) \geq P(Y \uplus Z)$. Since $T(X) = T(Y)$ and $T_g(X) = T_g(Y)$ for all g,

$$P(X) \geq P(Y) \iff \sum_{g \in \mathbf{G}} \sum_{n \in \mathbf{N}(X)} T^n \frac{\left(\frac{T_g^n}{T^n} - \frac{T_g(X)}{T(X)} \right)^2}{T(X) - T_g(X)} \geq \sum_{g \in \mathbf{G}} \sum_{m \in \mathbf{N}(Y)} T^m \frac{\left(\frac{T_g^m}{T^m} - \frac{T_g(Y)}{T(Y)} \right)^2}{T(Y) - T_g(Y)}$$

but $T_2^n = T^n - T_1^n$ and $T_2(X) = T(X) - T_1(X)$, so

$$\left(\frac{T_2^n}{T^n} - \frac{T_2(X)}{T(X)} \right)^2 = \left(\frac{T_1^n}{T^n} - \frac{T_1(X)}{T(X)} \right)^2$$

and similarly,

$$\left(\frac{T_2^m}{T^m} - \frac{T_2(Y)}{T(Y)} \right)^2 = \left(\frac{T_1^m}{T^m} - \frac{T_1(Y)}{T(Y)} \right)^2$$

so that

$$\begin{aligned} P(X) \quad \geq \quad P(Y) &\iff \sum_{n \in \mathbf{N}(X)} T^n \left(\frac{T_1^n}{T^n} - \frac{T_1(X)}{T(X)} \right)^2 \geq \sum_{m \in \mathbf{N}(Y)} T^m \left(\frac{T_1^m}{T^m} - \frac{T_1(X)}{T(X)} \right)^2 \\ &\iff \sum_{n \in \mathbf{N}(X)} T^n \left(\left(\frac{T_1^n}{T^n} \right)^2 - 2\frac{T_1^n}{T^n} \frac{T_1(X)}{T(X)} + \left(\frac{T_1(X)}{T(X)} \right)^2 \right) \\ &\geq \sum_{m \in \mathbf{N}(Y)} T^m \left(\left(\frac{T_1^m}{T^m} \right)^2 - 2\frac{T_1^m}{T^m} \frac{T_1(X)}{T(X)} + \left(\frac{T_1(X)}{T(X)} \right)^2 \right) \\ &\iff \sum_{n \in \mathbf{N}(X)} \frac{(T_1^n)^2}{T^n} \geq \sum_{m \in \mathbf{N}(Y)} \frac{(T_1^m)^2}{T^m} \end{aligned}$$

(In the first line we have eliminated the common factor $\frac{1}{T_1(X)} + \frac{1}{T_2(X)}$ and used the fact that $T_1(X) = T_1(Y)$ and $T(X) = T(Y)$.) A similar argument shows that

$$\begin{aligned} P(X \uplus Z) &\geq P(Y \uplus Z) \Longleftrightarrow \sum_{n\in\mathbf{N}(X)} T^n \left(\frac{T_1^n}{T^n} - \frac{T_1(Z \uplus X)}{T(Z \uplus X)}\right)^2 \geq \sum_{m\in\mathbf{N}(Y)} T^m \left(\frac{T_1^m}{T^m} - \frac{T_1(Z \uplus X)}{T(Z \uplus X)}\right)^2 \\ &\Longleftrightarrow \sum_{n\in\mathbf{N}(X)} \frac{(T_1^n)^2}{T^n} \geq \sum_{m\in\mathbf{N}(Y)} \frac{(T_1^m)^2}{T^m} \end{aligned}$$

so P satisfies IND1 in the case of two groups. IND1 is violated when there are more than two groups. A simple 3-group example suffices to show this. Let $X = \{(0, 2, 3), (6, 4, 3)\}$, $Y = \{(3, 2, 0), (3, 4, 6)\}$, and $Z = \{(0, 10, 100)\}$. Then $P(X) = P(Y)$ since P satisfies SYM, but one can verify that $P(X \uplus Z) \neq P(Y \uplus Z)$.

As for IND2, consider the districts $X = \langle(2, 0, 0), (0, 2, 0), (0, 0, 2)\rangle$, $Y = \langle(3, 0), (0, 3)\rangle$ and $Z = \langle(1, 0), (0, 1)\rangle$. It can be checked that $P(X \uplus Z) \neq P(Y \uplus Z)$ even though $P(X \uplus c(Z)) = P(Y \uplus c(Z))$, so P violates IND2. As for GDP, consider the district $X = \langle(0, 10), (10, 0)\rangle$ and the district $Y = \langle(0, 0, 5, 5), (5, 5, 0, 0)\rangle$ that is obtained after we split evenly each of X's groups. It can be checked that $P(X) \neq P(Y)$. Q.E.D.

Claim 6 *The index Cl violates SYM, SDP, IND2, SI, and SSD. It satisfies WSI, N, IND1, and C.*

Proof. SYM is clearly violated and WSI and N are obviously satisfied. As for SDP, suppose a district X contains a school n in which the proportion minority is exactly κ. Assume there is more than one minority student in n. Suppose a new school is built and a single minority student from n is moved to that school. The effect of this is to remove $T_2^n - 1$ minority students from the sum in the definition of the index. Hence, the index falls, violating SDP. Regarding IND1, let $X, Y \in \mathcal{C}$ be two districts with equal populations and equal group distributions. We must show that for any district Z that contains a single school, $Cl(X) \geq Cl(Y)$ if and only if $Cl(X \uplus Z) \geq Cl(Y \uplus Z)$. Since $T_2(X) = T_2(Y)$,

$$\begin{aligned} Cl(X) &\geq Cl(Y) \Longleftrightarrow \frac{1}{T_2(X)} \sum_{n\in\mathbf{N}(X):p_2^n\geq\kappa} T_2^n \geq \frac{1}{T_2(X)} \sum_{n\in\mathbf{N}(Y):p_2^n\geq\kappa} T_2^n \\ &\Longleftrightarrow \frac{1}{T_2(X) + T_2(Z)} \sum_{n\in\mathbf{N}(X\uplus Z):p_2^n\geq\kappa} T_2^n \geq \frac{1}{T_2(X) + T_2(Z)} \sum_{n\in\mathbf{N}(Y\uplus Z):p_2^n\geq\kappa} T_2^n \\ &\Longleftrightarrow Cl(X \uplus Z) \geq Cl(Y \uplus Z) \end{aligned}$$

so IND1 holds. As for IND2, say $\kappa = 0.5$, $X = \langle(1000, 500), (0, 500)\rangle$, $Y = \langle(10, 1)\rangle$, $Z = \langle(101, 0), (0, 100)\rangle$. Then $Cl(X \uplus Z) = 600/1100$, $Cl(Y \uplus Z) = 100/101$, $Cl(X \uplus c(Z)) = 500/1100$ and $Cl(Y \uplus c(Z)) = 0/101$, so IND2 is violated. SI is obviously violated. As for C, for any city W, let $a_\kappa(W)$ be the number of members of group 2 who attend schools in which at least a proportion κ are in group 2. Then for any $c \in [0, 1]$ and any districts X and Y, $a_\kappa(cX \uplus (1-c)Y) = ca_\kappa(X) + (1-c)a_\kappa(Y)$. Hence, $Cl(cX \uplus (1-c)Y) = \frac{ca_\kappa(X)+(1-c)a_\kappa(Y)}{cT_2(X)+(1-c)T_2(Y)}$ which is a continuous function of c. Hence C is satisfied. Q.E.D.

Claim 7 *The index CR violates SYM, SDP, IND1, IND2, SI, and SSD. It satisfies WSI, N, and C.*

Proof. SYM is clearly violated and WSI and N are clearly satisfied. SDP is violated: for the district $\langle(1,2,9)\rangle$, CR equals 0; if the school is split to yield $\langle(1,1,0),(0,1,9)\rangle$ CR falls to -0.2. IND1 is Violated: let $X = \langle(1,1,1),(1,0,1),(0,0,2)\rangle$, $Y = \langle(1,1,2),(1,0,2)\rangle$, and $Z = \langle(2,0,0)\rangle$. Then one can verify that $CR(X) > CR(Y)$ while $CR(X \uplus Z) < CR(Y \uplus Z)$. IND2 is violated: let $X = \langle(1,1,0),(0,0,98)\rangle$, $Y = \langle(30,20,0),(20,30,0)\rangle$, and $Z = \langle(10,0,0),(0,10,0)\rangle$. One can verify that $CR(X \uplus Z) > CR(Y \uplus Z)$ but $CR(X \uplus c(Z)) < CR(Y \uplus c(Z))$. SI is obviously violated. As for C, for any district W, let $a_1(W) = \sum_{n\in\mathbf{N}(W)} T_1^n \frac{T_2^n+T_3^n}{T^n}$ and $a_2(W) = \sum_{n\in\mathbf{N}(W)} T_2^n \frac{T_2^n+T_3^n}{T^n}$. Then for any $c \in [0,1]$ and any districts X and Y, $a_i(cX \uplus (1-c)Y) = ca_i(X) + (1-c)a_i(Y)$ for $i = 1,2$. Hence, $CR(cX \uplus (1-c)Y) = \frac{ca_2(X)+(1-c)a_2(Y)}{cT_2(X)+(1-c)T_2(Y)} - \frac{ca_1(X)+(1-c)a_1(Y)}{cT_1(X)+(1-c)T_1(Y)}$ which is a continuous function of c. Hence C holds. Q.E.D.