

More Places than Crimes: Implications for Evaluating the Law of Crime Concentration at Place

Wim Bernasco^{1,2}  · Wouter Steenbeek¹

Published online: 20 September 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract

Objectives The crime and place literature lacks a standard methodology for measuring and reporting crime concentration. We suggest that crime concentration be reported with the Lorenz curve and summarized with the Gini coefficient, and we propose generalized versions of the Lorenz curve and the Gini coefficient to correct for bias when crime data are sparse (i.e., fewer crimes than places).

Methods The proposed generalizations are based on the principle that the observed crime concentration should not be compared with perfect equality, but with maximal equality given the data. The generalizations asymptotically approach the original Lorenz curve and the original Gini coefficient as the number of crimes approaches the number of spatial units.

Results Using geocoded crime data on two types of crime in the city of The Hague, we show the differences between the original Lorenz curve and Gini coefficient and the generalized versions. We demonstrate that the generalizations provide a better representation of crime concentration in situations of sparse crime data, and that they improve comparisons of crime concentration if they are sparse.

Conclusions Researchers are advised to use the generalized versions of the Lorenz curve and the Gini coefficient when reporting and summarizing crime concentration at places. When places outnumber crimes, the generalized versions better represent the underlying processes of crime concentration than the original versions. The generalized Lorenz curve, the Gini coefficient and its variance are easy to compute.

✉ Wim Bernasco
wbernasco@nscr.nl

Wouter Steenbeek
wsteenbeek@nscr.nl

¹ Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), P.O. Box 71304, 1008 BH Amsterdam, The Netherlands

² Department of Spatial Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

Keywords Criminology of place · Crime concentration · Lorenz curve · Gini coefficient

Introduction

A defining feature of the criminology of place is its focus on the analysis of crime measured at micro-geographic units. Recently, Weisburd introduced the ‘law of crime concentration at place’. It states that “for a defined measure of crime at a specific microgeographic unit, the concentration of crime will fall within a narrow bandwidth of percentages for a defined cumulative proportion of crime.” (Weisburd 2015: 138). The formulation of the law is an important milestone in the evolution of the crime and place literature. It represents a strong claim regarding the ubiquity of crime concentration at places. To underline this, Weisburd (2015: 151) writes “The data suggest that the law of crime concentration is a “general proposition of universal validity” (Sutherland 1947: 23), analogous to physical laws observed in the natural sciences”.

Although crime concentration is one of the few stylized facts generally accepted in the criminology of place, the field has not yet developed common standards for reporting and summarizing crime concentrations. The law of crime concentration at place does not prescribe how crime concentration should be measured, although its wording suggests that it is best captured by relating cumulative percentages of places to cumulative percentages of crimes.

Indeed, most studies of crime concentration at places report their results using cumulative percentage statements of the form “Y percent of crime occurs in the X percent most targeted places”. However, there appears to be no accepted standard for choosing particular reference values of either X or Y, other than a slight tendency to fixate the value of Y at round numbers (different values of X and Y are chosen by, for example, Braga et al. 2010; Curman et al. 2015; Sherman et al. 1989; Weisburd 2015; Weisburd and Amram 2014). The lack of a common yardstick complicates comparisons between studies, and thus hampers testing the law of crime concentration. The literature on crime concentration at places also lacks a single measure of concentration to quantify the overall amount of concentration. Such a summary measure would be useful to test the law of crime concentration by comparing levels of crime concentration across areas, periods or crime types.

The first aim of this article is to argue why the *Lorenz curve* is an excellent candidate for a detailed and exhaustive description of crime concentration at places, and why the *Gini coefficient* is an excellent summary measure. The Lorenz curve is a graph that was originally designed to visualize income inequality, but has much broader applications. The Gini coefficient is closely related to the Lorenz curve, and summarizes the level of concentration in a single number between 0 and 1, the former representing a completely equal distribution of crimes across places, and the latter representing maximal concentration of all crimes in a single place. Both the Lorenz and the Gini are described in the next section.

However, as we will demonstrate, the Lorenz curve and the Gini coefficient *overestimate* the level of crime concentration if there are fewer crimes than places, a situation that is common in studies that use street segments or census blocks as the spatial unit of analysis (e.g., Andresen and Malleon 2011), and very common in studies that use individual addresses (e.g., Sherman et al. 1989). Because these micro-geographic units are small and crimes are relatively rare, crime data tend to be relatively sparse: the total number of crimes is usually smaller than the total number of micro-geographic units. As

crime is a discrete variable (a single crime cannot be divided over two places), in sparse data situations an equal distribution of crime across places is not possible. Drawing the Lorenz curve and calculating the Gini coefficient without accounting for this structural constraint hinders comparisons of crime concentration between cities, between crime types and between periods. As a consequence, it could incorrectly falsify or corroborate the law of crime concentration at places. If the Lorenz curve and the Gini coefficient are to be the preferred tools for describing and summarizing crime concentration at places, this bias must be addressed and corrected.

The second aim of this article is to provide such a correction. We propose a generalized way to display the Lorenz curve, and a generalization of the Gini coefficient. The generalized Lorenz curve and the generalized Gini coefficient asymptotically approach the traditional Lorenz curve and Gini coefficient when the number of crimes approaches and finally overtakes the number of places. The outcomes of the generalized Lorenz and Gini are thus comparable across areas, periods and crime types irrespective of whether the sample includes fewer or more crimes than units of analysis. To facilitate statistical tests, we also discuss a method of estimating the variance of the generalized Gini coefficient.

Although the article that inspired our work (Weisburd 2015) was not saddled with the issue raised here—all study regions in the analysis had much more crimes than street segments during the periods covered—future tests of the law of crime concentration at places are likely to include situations where crime data are sparse, as many studies start to focus on small spatial units of analysis, short time intervals, and specific crime types. Indeed, a number of key results in the literature on the criminology of place are based on sparse crime data. Braga et al. (2010) analyze the concentration of 7359 firearm incidents on 28,530 street units over a 29-year period. The authors point out that “The fact that each year, on average, there are fewer than 254 ABDW-Firearm incidents among nearly 28,530 street units suggests that even a purely random distribution might produce the observed clustering” (p. 42). Andresen and Malleon (2011) investigate 3 years of crime data for seven different crime types, analyzing the stability of the spatial patterns for 110 census tracts, 1011 dissemination areas, and 11,730 street segments. They report fewer assaults (in 1996 and 2001), theft (in 2001), robbery, sexual assault, and theft of vehicle (all years) than the total number of street segments. As a final example, Andresen et al. (2016) analyze eight crime types (assault, burglary, other, robbery, theft from vehicle, theft, theft of vehicle) over a 16-year period. Except for theft, theft from vehicle, and burglary in the early waves of data, the number of spatial units ($n = 18,445$) outnumber the number of crime incidents.

The issues we raise about concentration measures are not unique to the crime and place literature. They also apply to research on crime and victimization of individuals. In developmental and life-course criminology, it has been found that a relatively large share of crime is committed by a relative small share of individuals, and the phenomenon has been presented with Lorenz curves (Fox and Tracy 1988; van de Weijer et al. 2014). In line with a suggestion made by Fox and Tracy (1988), authors in these fields have generally selected from the data only those individuals that had offended at least once, and calculated the Lorenz curve and the Gini coefficient for this subsample. In victimology, concentration of victimization refers to the finding that a relatively large share of crime is committed against a relative small share of individuals. Tseloni and Pease (2005) demonstrated concentration of victimization amongst respondents in the British Crime Survey with Lorenz curves. They presented Lorenz curves for all respondents and also, analogously to what Fox and Tracy did for offending, for victims only. Thus, in both fields the Lorenz curve and the Gini coefficient have not only been used, but have also been adjusted to

satisfy certain research requirements. We suggest that the generalizations of the Lorenz curve and the Gini curve that we propose in this paper, could also be fruitfully applied in research on concentration in offending and victimization.

A concise roadmap closes this introduction. In the first section we briefly recapitulate the Lorenz curve and the Gini coefficient. The subsequent section demonstrates the limitations of the Lorenz and the Gini in situations where places outnumber crimes. The section that follows proposes the generalized Lorenz curve and Gini coefficient. The one that follows addresses statistical inference of the Gini, focusing on the estimation of its variance. Next, we demonstrate the virtues of the generalized methods on real empirical data of crime in street segments of the city of The Hague, the Netherlands. The final section summarizes what has been learned.

Lorenz Curve and Gini Coefficient

The Lorenz curve is a function that links the cumulative distribution of a variable (e.g. crime) to the cumulative distribution of observational units (e.g. places). The Lorenz curve has traditionally been used to visualize inequality in income distributions, but it has also been used in various other disciplines to visualize inequality or concentration (e.g., Damgaard and Weiner 2000). It can be used as a measure of concentration (or inequality) for any variable measured at ratio or interval level, in any sample or population.

Applied to the distribution of crimes across places, the Lorenz curve plots cumulative percentages of crime on the vertical axis against cumulative percentages of places on the horizontal axis, with the places ordered by number of crimes.¹ Thus, each and every point on the curve corresponds to a statement like “*Y* percent of crimes occur in the *X* percent most targeted places”. In other words, the Lorenz curve includes in a single graph *all* cumulative percentage statements that can be made about a given crime distribution! By presenting the Lorenz curve there is no longer any need to decide on a cutoff value for *X* or for *Y*, as all are included.

Quite surprisingly, amongst the dozens of studies that address crime concentration at places, to our knowledge only a few papers (Bowers 2014; Davies and Johnson 2015; Johnson 2010; Johnson and Bowers 2010; Steenbeek and Weisburd 2015) have used the Lorenz curve and the Gini coefficient.

For purposes of exposition, the Lorenz curve is best explained by presenting a simple, easily verifiable example with a small number of places. In Table 1 we present a fictitious distribution of 100 crimes across 10 places. The leftmost column lists the alphanumeric labels of the 10 places, while the second column lists the number of crimes in each place. The places are organized in descending order with respect to their number of crimes, starting with place G (22 events), followed by place A (18 events), place D (17 events) and so on, until place B (1 event) at the bottom.

The next two columns display cumulative percentages. The third column is the cumulative percentage of places: place G forms 10 percent of the places, places G and A together form 20 percent, places G, A and D together are 30 percent, and so on. The fourth column lists the cumulative percentages of crimes. A has 22 percent of crime, G and A

¹ In this paper, without loss of generality we order the units reversely, from high to low. This implies the Lorenz curve is completely located above rather than below the 45° reference line. The only reason for the reversal is that the resulting curve directly corresponds to statements like “*Y* percent of crimes occur in the *X* percent most targeted places”.

Table 1 Fictitious (cumulative) distribution of crime across places, rank order and intermediate quantities in Gini coefficient calculation

| Place label | Number of crimes | Cumulative percentage of places | Cumulative percentage of crimes | Proportion of crime y_i | Ranking (by y_i) i | $i \times y_i$ |
|-------------|------------------|---------------------------------|---------------------------------|---------------------------|-------------------------|----------------|
| G | 22 | 10 | 22 | 0.22 | 10 | 2.20 |
| A | 18 | 20 | 40 | 0.18 | 9 | 1.62 |
| D | 17 | 30 | 57 | 0.17 | 8 | 1.36 |
| J | 12 | 40 | 69 | 0.12 | 7 | 0.84 |
| C | 11 | 50 | 80 | 0.11 | 6 | 0.66 |
| E | 8 | 60 | 88 | 0.08 | 5 | 0.40 |
| J | 6 | 70 | 94 | 0.06 | 4 | 0.24 |
| F | 3 | 80 | 97 | 0.03 | 3 | 0.09 |
| H | 2 | 90 | 99 | 0.02 | 2 | 0.04 |
| B | 1 | 100 | 100 | 0.01 | 1 | 0.01 |
| Σ | 100 | | | | | 7.46 |

together have 40 percent of crime, and G, A and D together have 57 percent of crime. All 10 places together have a total of 100 percent of crime. The three rightmost columns are explained below and in Appendix 2 (“Expression of Adjusted Gini in Terms of Observed Cases”).

Figure 1 shows a Lorenz curve that plots the cumulative percentage of crimes against the cumulative percentage of places (i.e., the fourth against the third column of Table 1).

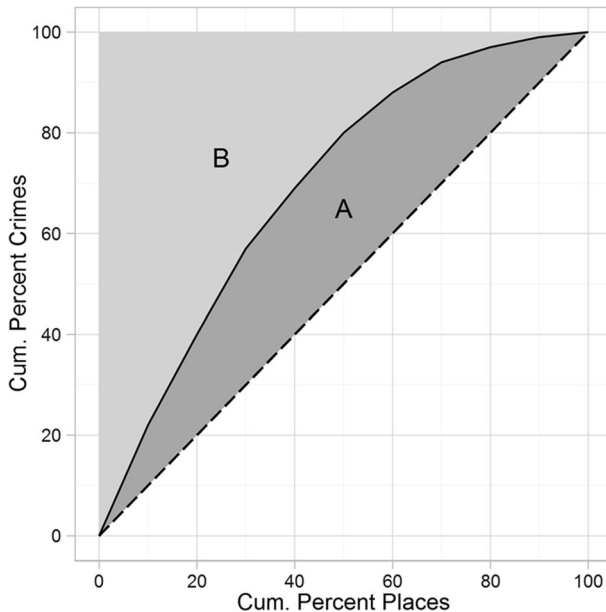


Fig. 1 Lorenz curve based on data in Table 1

The graph also includes a straight diagonal line with slope 1. This line represents *perfect equality*, a situation in which the distribution of crime cannot be made more equal by redistributing crimes from one place to another. If all places would have the same number of crimes (in the example each place would have exactly 10 crimes), the Lorenz curve would coincide with the line of perfect equality.

The Gini coefficient is a single number that quantifies the amount of concentration in a distribution. It was developed more than a century ago by Corrado Gini (1912) as a measure of income inequality.² The Gini coefficient can be geometrically defined as the ratio of the area between the Lorenz curve and the line of perfect equality, and the area above the line of perfect equality. In Fig. 1, we denote the area between the Lorenz curve and the line of perfect equality by A and the area above the Lorenz curve by B , so the Gini coefficient equals $A/(A + B)$. Theoretically, the Gini coefficient ranges between 0 and 1. If the Gini equals 0, the Lorenz curve coincides with the line of perfect equality, and crime concentration is minimal. If the Gini coefficient equals 1, crime is maximally concentrated and all crimes occur in a single place.³

The number of equivalent expressions for the Gini coefficient is remarkable large (Yitzhaki and Schechtman 2013). The most intuitive expression is the relative mean difference, which is the average of absolute differences in crimes between all possible pairs of places, divided by the mean number of crimes per place. For computational purposes, the following formula is more efficient:

$$G = \left(\frac{1}{n}\right) \left(2 \sum_{i=1}^n iy_i - n - 1\right) \quad (1)$$

where G is the Gini coefficient, n is the total number of places, y_i is the proportion of crimes occurring in place i , and i the rank order of the place when places are ordered by the number of crimes y . Using the calculation results in three rightmost columns of Table 1, it is easily verified that

$$G = \left(\frac{1}{10}\right) (2(7.46) - 10 - 1) = .392 \quad (2)$$

Limitation of Lorenz and Gini in Sparse Data Situations

In situations where the variable that is being studied is discrete and the total number of events is smaller than the number of units of analysis (e.g. if the total number of crimes is smaller than the number of places), the interpretation of the Lorenz curve and the Gini

² The Gini coefficient satisfies four preferred properties of inequality measures, namely anonymity (the identities of the places are irrelevant), scale independence (the overall level of crime is irrelevant), population independence (the total number of places is irrelevant) and the transfer principle (moving a crime from a higher-crime place to a lower-crime place reduces concentration). A major advantage of the Gini is its relation to the Lorenz curve, which makes it easier to interpret.

³ If the number of units is small, the Gini coefficient is biased because it is constrained to be lower than 1, a situation that can be corrected by multiplying the Gini coefficient by $n/(n-1)$, where n is the number of units in the sample (Deltas 2003). In this example, if all 100 crimes were concentrated in one of the 10 places, the biased value of the Gini coefficient would be .90, which would be corrected to 1 by multiplication with 10/9. In this article we will ignore this small-sample correction, as we are addressing situations in which this bias is negligible because the number of places is large.

coefficient are problematic. The reason for this is that both the interpretation of the Lorenz curve and the calculation of the Gini coefficient depend on the line of perfect equality. However, *if places outnumber crimes, perfect equality is not logically possible*. Because crimes are discrete and a single crime cannot be divided between places, some places necessarily have no crime, and thus less crime than other places. As a result, the boundary condition of maximal equality is no longer represented by a line with slope 1, but by a line with a steeper slope, which we denote as the *line of maximal equality*.

The two panels in Fig. 2 illustrate the issue. The left panel represents a situation in which there are $n = 10$ places and $c = 10$ crimes, and where $G = 0$. In other words, 10 crimes are equally distributed across 10 places, each of which has a single crime. The Lorenz curve coincides with the line of perfect equality.

Now imagine that we remove at random every other crime from the data. The result is the situation depicted in the right panel of Fig. 2, where $n = 10$ places and $c = 5$. Because 5 crimes cannot be distributed equally over 10 places, half the places have 1 crime and the other half have no crimes. The resulting Gini coefficient equals .50. This outcome is problematic, because without any change in an underlying distributional mechanism (we have merely taken out half the crimes at random), the concentration of crime at place has substantially increased.

Note that the change affects both the Gini coefficient (which changes from 0 to .50) and the presentation of the Lorenz curve and the cumulative percentage statements it represents. For example, in the left panel of Fig. 2 “40 percent of crime occurs in 40 percent of places” whereas in the right panel of Fig. 2 “80 percent of crime occurs in 40 percent of places”. The latter statement might better be replaced by: “40 percent of crime occurs in 40 percent of the places where crime could occur.” This limitation was already identified in the article that launched the criminology of place (Sherman et al., 1989), in which the authors analyzed calls for service originating from individual addresses in Minneapolis. They demonstrated a strong concentration of calls for service from particular addresses:

“Taken separately, each of the predatory crimes shows even greater geographic concentration than all calls for service: all 4166 robbery calls were located at only 2.2 % (against a possible 3.6 %) of all places, all 3908 auto thefts at 2.7 % (against a possible 3.4 %), and all 1729 rape/CSCs at just 1.2 % (against a possible 1.5 %) of the places in the city.” (Sherman et al. 1989: 39).

Note that the authors compare the observed levels of concentration (e.g. 100 percent of robbery calls originated from 2.2 percent of the addresses) with the level of concentration that *would be observed if crime were maximally dispersed*: if each call came from a different address, 100 percent of robbery calls would have originated from 3.6 percent of the addresses.⁴ In other words: the 3.6 percent puts the 2.2 percent in another perspective that makes it much less spectacular.

Although the Lorenz curves, cumulative percentage statements, and Gini coefficients in Fig. 2 correctly describe the observed amounts of concentration in both samples (crime is more concentrated in the right panel than in the left panel) the interpretation in terms of an underlying stochastic process is problematic: the higher concentration of crime in the right panel is an artifact of the sparsity and discreteness of the data, i.e. of $n > c$ and $y_i \in \mathbb{N}$.

⁴ The authors did not make this calculation completely explicit. They estimated the total number of places in the city at 115,000 (p. 37), so that the 4166 robberies could have occurred at no more than $100 \times (4166/115,000) = 3.6$ percent of the places.

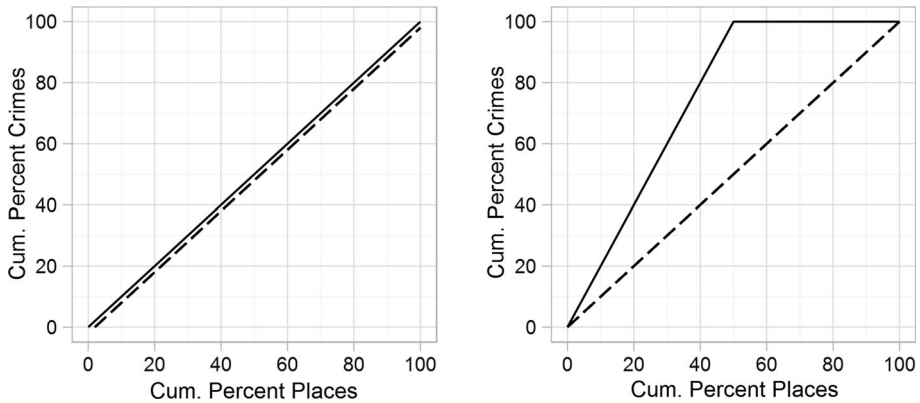


Fig. 2 Lorenz curves (*left*: 10 places, 10 crimes, $G = 0$, *right*: 10 places, 5 crimes, $G = .5$)

Generalized Lorenz Visualization and Gini Calculation

To address the shortcomings discussed in the previous section, we generalize the Gini coefficient and the visualization of the Lorenz curve. Our solution does not change the Lorenz curve itself, but replaces the line of perfect equality with a line of maximal equality. The rationale is that this line is a reference line and should represent the boundary condition of maximal equality. It should therefore only have a slope of 1 if the total number of crimes c is larger than or equal to the total number of places n . In all other cases its slope should be $\frac{n}{c}$, the ratio of the number of places (n) and the number of crimes (c). In other words, the slope of the line of maximal equality should be

$$\frac{n}{\min(c, n)} = \max\left(\frac{n}{c}, 1\right) \tag{3}$$

The simple example provided in Fig. 3 will help us explain the required generalizations. The figure is based on a situation where there are 10 places: one place has 3 crimes, two other places have 1 crime, and the other five places have 0 crimes. The Lorenz curve (solid line) does not behave differently from the previous figures: it plots the cumulative percentage of crimes for each cumulative percentage of places ordered by crime frequency. What is different is the line of maximal equality (dotted line), which starts in the origin but now has a slope of $\frac{n}{c} = \frac{10}{5} = 2$. This visualization improves on the original visualization in which only the Lorenz curve and the line of perfect equality (dashed line) are presented, because the original visualization hides the fact that the boundary condition is at the line of maximal quality (dotted) and not at the line of perfect equality (dashed). If the number of crimes $c < n$, the generalized line of maximal equality lies completely above the line of perfect equality. When $c \geq n$, the two coincide. Thus, the Lorenz curve itself does not change in case of sparse data. The innovation is the line of maximal equality.

The generalized Gini we propose takes into account that a Gini value of 0, perfect equality, cannot be attained if $c < n$. In order to correct for this property, the generalized Gini uses the line of maximal equality. The generalized Gini is defined analogously to the original Gini, but with reference to the line of maximal equality rather than the line of

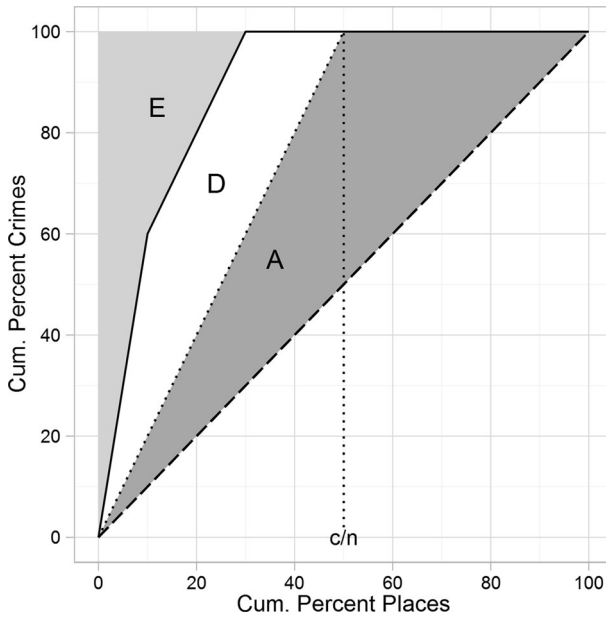


Fig. 3 Generalized Lorenz curve, 10 places, 3 + 1 + 1 crimes, $G = .78$, $G' = .58$

perfect equality. Thus, whereas the original Gini coefficient in Fig. 3 equals $(A + D)/(A + D + E)$, when $c < n$, the generalized Gini equals $D/(D + E)$.⁵

The generalized Gini coefficient can be expressed as a function of c , n , and the original Gini coefficient G as follows (proof in Appendix 1, “Expression of Adjusted Gini in Terms of Original Gini”).

$$G' = \max\left(\frac{n}{c}, 1\right)(G - 1) + 1 \tag{4}$$

To facilitate implementation in software, the generalized Gini coefficient can also be expressed in terms of individual observations, without reference to the original Gini coefficient, as follows (proof in Appendix 2, “Expression of Adjusted Gini in terms of observed cases”).

$$G' = \max\left(\frac{1}{c}, \frac{1}{n}\right) \left(2 \sum_{i=1}^n (iy_i) - n - 1 \right) - \max\left(\frac{n}{c}, 1\right) + 1 \tag{5}$$

In all situations where $c < n$, the generalized Gini coefficient makes a downward correction to the original Gini. For example, in Fig. 3 the original Gini coefficient equals .78, but the generalized Gini only .58. If $c \geq n$, the generalized Gini coefficient G' equals the original Gini coefficient G .

⁵ Fox and Tracy (1988) propose an alternative summary measure α that is similar to our adjusted Gini measure. Rather than use c/n to define the point where the line of maximal equality reaches $Y = 1$, they use q/n to define that point, where q is the number individuals committing 1 or more offenses. We prefer our own measure because it does not exclude places with zero crimes automatically. Our measure only excludes places with zero crimes if they are imposed by sparse crime data.

The use of a simulated reference curve (Davies and Johnson 2015; Johnson 2010; Johnson and Bowers 2010) provides another solution to the issue of sparse crime data. Johnson and his coauthors repeatedly (99 times) assigned c crimes to n places at random⁶ and calculated a Lorenz curve on these simulated data. They subsequently used the means across all runs of the simulation to calculate the reference curve for the real data. Clearly, when there are more places than crimes, in each run of the simulation the resulting Lorenz curve must necessarily lie above the line of perfect equality, and so must the average curve. In other words, because the simulation method is based on permutations of the observed (sparse) data, the reference curve it creates automatically accounts for the structural impossibility of perfect equality in the distribution of crime. The same holds for their Gini coefficient, as it is defined as the area between the observed Lorenz curve and the average reference line across all simulation runs.

While simulation is a powerful and flexible method to generate counterfactual distributions, we believe our analytical solution is to be preferred not only because it is easier to implement, but also because it is a simple generalization of ideas, methods and techniques that have proven their value for more than a century.

Statistical Inference: Testing the Law of Crime Concentration

Almost all research on crime concentration uses population data rather than samples. They include all street segments, all addresses, or all restaurants in a given catchment area. Sample and population coincide and, strictly speaking, statistical inference is not necessary. In these cases, testing the law of crime concentration boils down to comparing absolute levels of concentration qualitatively (as in Weisburd 2015).

In other situations researchers may have a sample from a larger population, and need inferential statistics. Langel and Tillé (2013) provide a useful overview of variance estimation and statistical tests of the Gini coefficient. In situations where sampling structures are simple (e.g. plain random sampling), the jackknife re-sampling approach (Efron and Stein 1981) seems to best fulfill the need of most studies of crime concentration. The jackknife variance estimator of the Gini coefficient is

$$\text{Var}(G) = \left(\frac{n-1}{n}\right) \sum_{k=1}^n (G(n, k) - \bar{G}(n))^2 \quad (6)$$

where G is the Gini coefficient, n is the total number of places, $G(n, k)$ is the Gini coefficient calculated on the $(n-1)$ places after removing the k th place, and $\bar{G}(n)$ is mean of all $G(n, k)$, $k = 1, 2, \dots, n$ (Ogwang 2000). By replacing G by G' in the above equation, we obtain the jackknife estimator of the variance of the generalized Gini coefficient.

Illustration Using Sparse Crime Data

This section demonstrates the use of the generalized Lorenz curves and Gini coefficients using police recorded crimes in the city of The Hague, the Netherlands. Each crime is provided with coordinates of its location (see Steenbeek and Weisburd 2015 for details).

⁶ The issue they solve is actually slightly more complex because they randomly assign burglaries to residential units and subsequently aggregate the results to street segments.

For the purposes of this study these crime locations are assigned to the nearest street segment. The Hague consists of 14,375 street segments. They have a mean length of about 94 meters (310 feet) and a standard deviation of 108 meters (353 feet).

The data indeed show that situations of sparse crime data—in which the number of spatial units is greater than the number of crimes—are common. In the period 2007–2009, the police registered 1388 auto thefts, 430 sexual offenses, 1881 street robberies, 7135 assaults and 37,121 cases of break and enter. Break and enter combines all attempted or completed thefts committed by breaking into and entering a closed structure without permission and with the intent to steal or commit another crime. It includes residential burglary, commercial burglary, and theft from caravans, ships or automobiles.

As the purpose of the present study is to demonstrate that the generalized Gini coefficient provides a better estimate of the crime concentrations in situations of sparse crime data, we purposefully selected a crime type where $c > n$ (break and enter) and compare it to one where $c < n$ (assault). Three years of data (2007–2009) were summed in order to have enough cases per crime type to make for an interesting example.⁷

We first present descriptive statistics on the percentage of street segments that accounts for 50 percent of all crime. For both types of crime, a very small percentage of street segments account for the bulk of the crime: choosing $Y = 50$, about 8.84 percent of the street segments account for half of all break and enter cases, and about 4.70 percent of the street segments account for half of all assaults. Choosing $X = 10$, another way to summarize the crime concentration is by stating that for break and enter and assault, the percentage of crimes committed in the 10 percent most victimized street segments, is 53 and 73 percent, respectively. Obviously, both crime types are highly concentrated in a small number of places.

The Lorenz curve is an improvement over such cumulative percentage statements, as no arbitrary cut-off percentages need be chosen. Figure 4 shows the two Lorenz curves and their corresponding Gini coefficients. The figure shows that assault is highly concentrated, but that break and enter is concentrated as well—both curves are located far above the line of maximal equality (the diagonal). The difference in concentration is reflected in their Gini coefficients of 0.73 for break and enter, and 0.86 for assault. The two Lorenz curves do not cross, implying that for *every* X percent of spatial units, assault is more concentrated than break and enter, a fact that cannot be gleaned from the descriptive cumulative percentages alone.

We next turn to the generalized Lorenz curve and generalized Gini coefficient for both crime types, presented in Fig. 5. For break and enter, the number of crimes ($c = 37,121$) is much larger than the number of street segments ($n = 14,375$). Therefore the line of maximal equality and the Gini coefficient for *break and enter do not differ* from the line of perfect equality and Gini coefficient: 0.73. However, the number of assaults ($n = 7135$) is smaller than the number of street segments and therefore the generalized Gini coefficient differs from the original. Figure 5 shows the generalized lines of maximal equality and the original (G) and generalized (G') Gini coefficients for break and enter and for assault. While all assaults over the 2007–2009 period in The Hague are concentrated in only 23 percent of all street segments, this is *against a possible* 50 percent of all street segments. This limit is reflected in the generalized line of maximal equality and the generalized Gini coefficient of

⁷ If we had only selected a single year, even the number of break and enter incidents would have been smaller than the number of street segments, i.e. lower than 14,375.

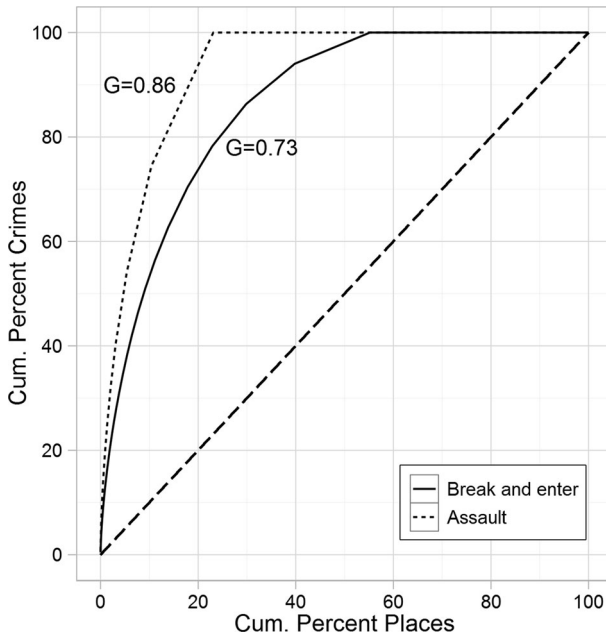


Fig. 4 Lorenz curves for burglary and assault, The Hague, 2007–2009

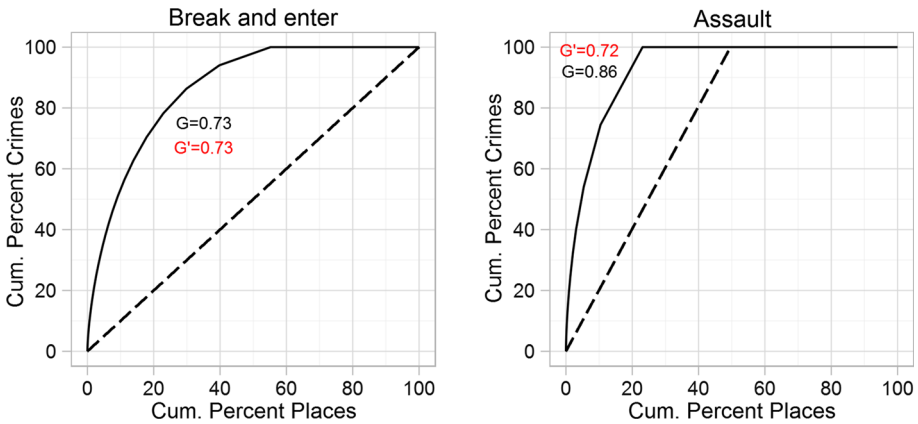


Fig. 5 Original and generalized Lorenz curves and Gini coefficients for break and enter and assault, The Hague, 2007–2009

$$G' = \max\left(\frac{n}{c}, 1\right)(G - 1) + 1 = \left(\frac{14,375}{7,135}\right)(0.86 - 1) + 1 = 0.72 \tag{7}$$

Thus, the generalized Gini coefficients of break and enter and of assault are approximately equal (.73 vs. .72), whereas the unadjusted coefficients are markedly different (.73 vs. .86 respectively) In summary, a traditional unadjusted comparison of break and enter and assault concentrations would have let us to believe that assault is more concentrated than break and enter, as implied by Fig. 4 and by .86 being larger than .73. However,

although assault is more concentrated than break and enter, the possible range of concentration for assault is much smaller. The rareness of the crime makes assault seem to be more concentrated than it actually is. Adjusting for this structural constraint, the generalized Gini coefficients show that the concentrations of assault and break and enter are actually remarkably similar.

Discussion

This article suggested that proper tests of the law of crime concentration require standards for measuring and reporting crime concentration. The Lorenz curve and the Gini coefficient were introduced as likely candidates. We explained why situations where places outnumber crimes can lead to flawed tests of the law of crime concentration, not only when the Lorenz and the Gini are used, but also when concentration is described conventionally, using cumulative percentage statements of the form “ Y percent of crime takes place in the X percent most targeted places”. We demonstrated how the issue can be solved transparently and elegantly by not changing the Lorenz curve itself but only its reference, and by using the generalized Gini coefficient. Our empirical demonstration suggested that in the calculation and the comparison of crime concentration measures, the generalized Lorenz curve and Gini coefficient serve two important functions. First, they add nuance to the significance of observed levels of concentration, which in case of sparse data must necessarily be high. Second, they provide better opportunities to compare crime concentrations over different crime types, periods and areas.

A minor caveat of the generalized Lorenz curve is, that it loses most of its function as a visualization tool in situations where c is much smaller than n . If, for example, $n = 20c$, the line of maximal equality reaches its 100 percent ceiling already at 5 percent. If this is the case (see, for example, the crime concentration percentages of specific crime types in Sherman et al. 1989) the area between the Lorenz curve and the line of maximal equality becomes very thin and will be difficult to see. The generalized Gini coefficient, however, remains useful as an overall quantification measure. A possible solution to such situations could be to simply rescale the x-axis, so that the maximum equals c/n rather than 1. This solution makes it easier to derive percentage statements from the graph, but makes comparisons difficult because they require common scales on the axes.

The measures of crime concentration at places that we propose in this article are primarily meant to facilitate testing Weisburd’s law of crime concentration at places. They may also help practitioners better understand crime concentrations: when there are more places than crimes, the adjusted Lorenz curve and Gini coefficient help to separate the volume of crime from the relative concentration of crime across units. However, we realize that practitioners usually *are* interested in the volume of crime: ‘hotspots’ guide police resource allocation as it helps them allocate manpower to where and when it is needed most. In operational decisions, plain numbers of crimes potentially prevented or solved are more informative and more useful than relative measures, because the latter standardize on potential targets and thus hide differences in the volume of crime.

In conclusion, we suggest that using the generalized Lorenz visualization and Gini calculation could prevent the law of crime concentration from being incorrectly falsified or corroborated. To emphasize this point, we suggest the law could be reformulated as follows:

For a defined measure of crime at a specific micro-geographic unit, the observed concentration of crime, *relative to its concentration under maximally possible dispersion*, will fall within a narrow bandwidth of percentages for a defined cumulative percentage of crime.

Acknowledgments We thank the unit The Hague of the Dutch National Police for providing crime data, and Arjan Blokland, Henk Elffers, Stijn Ruiter, Dirk Sierag, JQC Associate Editor Anthony Braga and three anonymous reviewers for suggestions that helped us improve the paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendices

Appendix 1: Expression of Adjusted Gini in Terms of Original Gini

We prove Eq. (4), stating that that, for $c > 0$ and $n > 0$, the Gini coefficient G' can be written as a function of the original Gini coefficient G , the number of places n , and the number of crimes c , as follows

$$G' = \max\left(\frac{n}{c}, 1\right)(G - 1) + 1$$

Both G and G' are defined as ratios of the areas A, D and E in Fig. 3.

$$G = \frac{A + D}{A + D + E} \quad (8)$$

$$G' = \frac{D}{D + E} \quad (9)$$

Because the 45° diagonal line of maximal equality has slope 1,

$$A + D + E = \frac{1}{2} \quad (10)$$

Substituting the left-hand side of (10) in (8) and rearranging terms yields

$$D = \frac{1}{2}G - A \quad (11)$$

Areas D and E together form a right-angled triangle with height 1 and base c/n , so that

$$D + E = \frac{c}{2n} \quad (12)$$

Substituting (12) in the numerator and (11) in the denominator of (9) yields

$$G' = \frac{\frac{1}{2}G - A}{c/2n} \quad (13)$$

Substitution of (12) in (10) yields

$$A = \frac{1}{2} - \frac{c}{2n} \tag{14}$$

Finally, substituting (14) in (13) and rearranging terms yields

$$G' = \frac{n}{c}(G - 1) + 1 \tag{15}$$

Note that Eq. (15) holds if $0 < c < n$, while $G' = G$ if $c \geq n$. Therefore, in the general case we have

$$G' = \max\left(\frac{n}{c}, 1\right)(G - 1) + 1 \tag{16}$$

This completes the proof.

Appendix 2: Expression of Adjusted Gini in Terms of Observed Cases

We prove Eq. (5), stating that, for $c > 0$ and $n > 0$, the adjusted Gini coefficient can be expressed as

$$G' = \max\left(\frac{1}{c}, \frac{1}{n}\right) \left(2 \sum_{i=1}^n (iy_i) - n - 1 \right) - \max\left(\frac{n}{c}, 1\right) + 1$$

where y_i is the proportion of all crime occurring in place i . The proof is partial, as we do not derive the equation from the geometry in Fig. 3, but start by quoting a common expression for the original Gini coefficient (e.g., Damgaard and Weiner 2000)

$$G = \frac{\sum_{i=1}^n (2i - n - 1)x_i}{n^2\mu} \tag{17}$$

in which; i = rank order index of the place when sorted in decreasing order with respect to y ; x_i = number of crimes in the i^{th} place; μ = the mean number of crimes across all n places, i.e.

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \tag{18}$$

Substitution of (18) in (17) yields

$$G = \frac{\sum_{i=1}^n (2i - n - 1)x_i}{n^2 \frac{1}{n} \sum_{i=1}^n x_i} \tag{19}$$

We define y_i as the proportion of crime occurring in place i :

$$\begin{aligned} y_i &\equiv \frac{x_i}{n} \\ x_i &= ny_i \end{aligned} \tag{20}$$

Substitution of (20) into (19) leads to

$$G = \left(\frac{1}{n}\right) \sum_{i=1}^n (2i - n - 1)y_i \tag{21}$$

Rearrangement of terms yields:

$$G = \left(\frac{1}{n}\right) \left(2 \sum_{i=1}^n (iy_i) - n - 1\right) \quad (22)$$

In Appendix 1 it was proven in Eq. (16) that

$$G' = \max\left(\frac{n}{c}, 1\right) (G - 1) + 1 \quad (16)$$

Substituting (22) into (16) yields

$$G' = \max\left(\frac{n}{c}, 1\right) \left(\left(\frac{1}{n}\right) \left(2 \sum_{i=1}^n (iy_i) - n - 1\right) - 1\right) + 1 \quad (23)$$

Rearrangement of terms leads to

$$G' = \max\left(\frac{n}{c}, 1\right) \left(\left(\frac{1}{n}\right) \left(2 \sum_{i=1}^n (iy_i) - n - 1\right)\right) - \max\left(\frac{n}{c}, 1\right) + 1 \quad (24)$$

$$G' = \max\left(\frac{1}{c}, \frac{1}{n}\right) \left(2 \sum_{i=1}^n (iy_i) - n - 1\right) - \max\left(\frac{n}{c}, 1\right) + 1 \quad (25)$$

This completes the proof.

References

- Andresen MA, Malleon N (2011) Testing the stability of crime patterns: implications for theory and policy. *J Res Crime Delinq* 48:58–82. doi:[10.1177/0022427810384136](https://doi.org/10.1177/0022427810384136)
- Andresen MA, Curman AS, Linning SJ (2016) The trajectories of crime at places: understanding the patterns of disaggregated crime types. *J Quant Criminol*. doi:[10.1007/s10940-016-9301-1](https://doi.org/10.1007/s10940-016-9301-1)
- Bowers K (2014) Risky facilities: crime radiators or crime absorbers? a comparison of internal and external levels of theft. *J Quant Criminol* 30:389–414. doi:[10.1007/s10940-013-9208-z](https://doi.org/10.1007/s10940-013-9208-z)
- Braga AA, Papachristos AV, Hureau DM (2010) The concentration and stability of gun violence at micro places in boston, 1980–2008. *J Quant Criminol* 26:33–53. doi:[10.1007/s10940-009-9082-x](https://doi.org/10.1007/s10940-009-9082-x)
- Curman ASN, Andresen MA, Brantingham PJ (2015) Crime and place: a longitudinal examination of street segment patterns in vancouver, BC. *J Quant Criminol* 31:127–147. doi:[10.1007/s10940-014-9228-3](https://doi.org/10.1007/s10940-014-9228-3)
- Damgaard C, Weiner J (2000) Describing inequality in plant size or fecundity. *Ecology* 81:1139–1142. doi:[10.1890/0012-9658\(2000\)081\[1139:DIIPSO\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[1139:DIIPSO]2.0.CO;2)
- Davies T, Johnson SD (2015) Examining the relationship between road structure and burglary risk via quantitative network analysis. *J Quant Criminol* 31:481–507. doi:[10.1007/s10940-014-9235-4](https://doi.org/10.1007/s10940-014-9235-4)
- Deltas G (2003) The small-sample bias of the Gini coefficient: results and implications for empirical research. *Rev Econ Stat* 85:226–234. doi:[10.1162/rest.2003.85.1.226](https://doi.org/10.1162/rest.2003.85.1.226)
- Efron B, Stein C (1981) The jackknife estimate of variance. *Ann Stat* 9:586–596
- Fox JA, Tracy PE (1988) A measure of skewness in offense distributions. *J Quant Criminol* 4:259–274
- Gini CW (1912) Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche [Variability and mutability: contribution to the study of statistical distributions and relations]: Tipogr. di P. Cuppini
- Johnson SD (2010) A brief history of the analysis of crime concentration. *Eur J Appl Math* 21:349–370. doi:[10.1017/S0956792510000082](https://doi.org/10.1017/S0956792510000082)
- Johnson SD, Bowers KJ (2010) Permeability and burglary risk: are Cul-de-Sacs safer? *J Quant Criminol* 26:89–111. doi:[10.1007/s10940-009-9084-8](https://doi.org/10.1007/s10940-009-9084-8)

- Langel M, Tillé Y (2013) Variance estimation of the Gini index: revisiting a result several times published. *J R Stat Soc: Ser A (Stat Soc)* 176:521–540. doi:[10.1111/j.1467-985X.2012.01048.x](https://doi.org/10.1111/j.1467-985X.2012.01048.x)
- Ogwang T (2000) A convenient method of computing the Gini index and its standard error. *Oxford Bull Econ Stat* 62:123–129. doi:[10.1111/1468-0084.00164](https://doi.org/10.1111/1468-0084.00164)
- Sherman L, Gartin PR, Buerger ME (1989) Hot spots of predatory crime: routine activities and the criminology of place. *Criminology* 27:27–55
- Steenbeek W, Weisburd D (2015) Where the action is in crime? an examination of variability of crime across different spatial units in The Hague, 2001–2009. *J Quant Criminol* 32:449–469. doi:[10.1007/s10940-015-9276-3](https://doi.org/10.1007/s10940-015-9276-3)
- Tseloni A, Pease K (2005) Population inequality: the case of repeat crime victimization. *Int Rev Victimol* 12:75–90. doi:[10.1177/026975800501200105](https://doi.org/10.1177/026975800501200105)
- van de Weijer SGA, Bijleveld CCJH, Blokland AAJ (2014) The intergenerational transmission of violent offending. *J Fam Violence* 29:109–118. doi:[10.1007/s10896-013-9565-2](https://doi.org/10.1007/s10896-013-9565-2)
- Weisburd D (2015) The law of crime concentration and the criminology of place. *Criminology* 53:133–157. doi:[10.1111/1745-9125.12070](https://doi.org/10.1111/1745-9125.12070)
- Weisburd D, Amram S (2014) The law of concentrations of crime at place: the case of Tel Aviv-Jaffa. *Police Pract Res* 15:101–114. doi:[10.1080/15614263.2013.874169](https://doi.org/10.1080/15614263.2013.874169)
- Yitzhaki S, Schechtman E (eds) (2013) More than a dozen alternative ways of spelling Gini. In: *The Gini methodology: a primer on statistical methodology*. Springer, New York, pp 11–31. doi:[10.1007/978-1-4614-4720-7_2](https://doi.org/10.1007/978-1-4614-4720-7_2)