# Measuring *Marginal* Crime Concentration: A New Solution to an Old Problem *


Aaron Chalfin
University of Pennsylvania

Jacob Kaplan
University of Pennsylvania

Maria Cuellar
University of Pennsylvania



**Abstract**

**Objectives:** In his 2014 Sutherland address to the American Society of Criminology, David Weisburd demonstrated that the share of crime that is accounted for by the most crime-ridden street segments is notably high and strikingly similar across cities, an empirical regularity referred to as the "law of crime concentration." In the large literature that has since proliferated, there remains considerable debate as to how crime concentration should be measured empirically. We suggest a measure of crime concentration that is simple, accurate and easily interpreted.

**Methods:** Using data from three of the largest cities in the United States, we compare observed crime concentration to a counterfactual distribution of crimes generated by randomizing crimes to street segments. We show that this method avoids a key pitfall that causes a popular method of measuring crime concentration to considerably overstate the degree of crime concentration in a city.

**Results:** While crime is significantly concentrated in a statistical sense and while some crimes are substantively concentrated among hot spots, the precise relationship is considerably weaker than has been documented in the empirical literature.

**Conclusions:** The method we propose is simple and easily interpretable and compliments recent advances which use the Gini coefficient to measure crime concentration.



*Keywords*: Criminology of place, hot spots, microgeography, Law of crime concentration



*We thank Purvasha Chakravarti, John MacDonald, Brendon McConnell, Dan OBrien, Alex Reinhart and Mauricio Sadinle for helpful comments on a previous version of this manuscript. Of course, all remaining errors are our own.


Corresponding author: Aaron Chalfin, Department of Criminology, 558 McNeil Building, University of Pennsylvania, E-Mail: achalfin [at] sas.upenn.edu.

## 1 Introduction

A large and growing literature in criminology documents the importance of place — in particular, microgeographic places like street segments — one of the two faces of a standard city block — in explaining crime. Across a large number of places and in a variety of contexts, crime is found to be highly concentrated (Sherman et al., 1989; Eck et al., 2007; Weisburd, 2015; Andresen et al., 2017; Haberman et al., 2017) and persistent over time (Weisburd et al., 2009; Gorr and Lee, 2015). Taken as a whole, the substantial geographic concentration of crime suggests that the social and physical features of the urban landscape might potentially play an important role in the crime production function and therefore that crime hot spots are an appropriate target over which a social planner can focus resources and ultimately intervene.[1] However, the efficiency with which resources can be targeted to crime hot spots depends critically on the extent to which crime is, in fact, concentrated. Consequently, an over-emphasis on place may crowd out other promising approaches to crime control (e.g., social service-based strategies) if the evidence on spatial crime concentration is misleading.

In his 2014 Edmund H. Sutherland address to the American Society of Criminology, David Weisburd summarized the research on the importance of place and noted that places have been studied far less by criminologists than other natural units of analysis (Weisburd, 2015). Weisburd further notes the extent to which crime is concentrated among the most crime-ridden street segments is remarkably consistent across cities and proposes that this empirical regularity is sufficiently strong to be characterized as a "law of crime concentration."[2] Across eight cities of varying sizes, the top one percent of street segments, ranked by crime incidence, accounted for approximately 25 percent of crimes in that city and the top 5 percent of street segments accounted for half of the crimes. The stability of these estimates is noteworthy and forms the basis for the claim that this pattern can be characterized as a law.

Despite the abundance of research inspired by the law of crime concentration, recent scholarship has raised a number of key measurement issues in how crime concentration should actually be measured

[1]Indeed, the empirical regularity that crime is highly spatially concentrated has been central to the study of criminal justice policy and has promulgated a number of important research literatures that have become a mainstay of empirical criminology including a large literature on hot spots policing (Weisburd and Green, 1995; Sherman and Weisburd, 1995; Braga, 2001; Braga and Bond, 2008; Weisburd and Telep, 2014; Braga et al., 2014) and the equally important literature on the importance of environmental design including research on restoring vacant lots (Branas et al., 2011; Garvin et al., 2013; Bogar and Beyer, 2016; Kondo et al., 2016; Branas et al., 2018; South et al., 2018; Moyer et al., 2019), reducing physical disorder (Kelling et al., 1982; Keizer et al., 2008; Skogan, 2012; Braga et al., 2015) and improving ambient lighting (Farrington and Welsh, 2002; Welsh and Farrington, 2008; Doleac and Sanders, 2015; Chalfin et al., 2019, 2020).

[2]In Weisburd's own words, "for a defined measure of crime at a specific microgeographic unit, the concentration of crime will fall within a narrow bandwidth of percentages."

(Bernasco and Steenbeek, 2017; Hipp and Kim, 2017; Levin et al., 2017; Prieto Curiel, 2019; OBrien, 2019; Mohler et al., 2019). In particular, prior research notes that the fact that a small share of street segments accounts for a large share of the crime over a given time period does not necessarily mean that crime is substantively concentrated. To see this, consider that even in the cities with most the challenging crime problems, the number of street segments can far exceed the number of crimes known to law enforcement over any reasonable time window. For instance, consider a city like New York in which there are approximately 120,000 street segments and 300 homicides annually. In this case, it is trivial to see that, even if each homicide occurs on a different street segment (thus, by definition, there would be no concentration of crime), 0.25 percent of the street segments would account for 100 percent of the homicides.[3] Thus, using the standard metric of crime concentration, the extent to which at least some types of crimes are concentrated will be biased upward. Similarly, the standard metric does not allow for a comparative analysis of concentration among different types of crimes since rarer crimes will, for mechanical reasons, appear to be more concentrated than more common crimes (Hipp and Kim, 2017).

Recent scholarship has proposed several modifications to the measurement of crime concentration that address these concerns (Bernasco and Steenbeek, 2017; Hipp and Kim, 2017; Levin et al., 2017; Curiel et al., 2018; Mohler et al., 2019; OBrien, 2019). A particularly common approach that is advanced by Levin et al. (2017) and which can be found in abundance in the recent literature (see e.g., Steenbeek and Weisburd (2016), Andresen et al. (2017), Schnell et al. (2017) and Umar et al. (2020)) is to measure crime concentration *only among street segments that experienced at least one crime*. The idea behind this approach is that crimes can only be concentrated where they, in fact, occur. This modification to the measurement of crime concentration does tend to reduce the degree of the bias in the standard measure but, as we show, in most empirical applications, removing crime-free street segments will continue to lead to a substantial overestimate of the extent to which crimes are concentrated.

In this article, we propose a different way to measure crime concentration that is simple, easily interpreted and which fully addresses the concerns outlined above. In particular, we compare actual concentration — for instance, the share of street segments accounting for 25 percent or 50 percent of the crimes — to a counterfactual level of crime concentration that is constructed by randomly assigning crimes to street segments, with replacement. Using randomization we generate a spatial distribution of crime where crime is

[3]Even over a period of ten such years, if every homicide occurred on a different street segment, we would observe that just 2.5 percent of the street segments account for 100 percent of the homicides.

not concentrated by construction.[4] Notably, our method compliments an alternative and highly convenient method of measuring crime concentration using the Gini coefficient (Bowers, 2014; Davies and Johnson, 2015; Steenbeek and Weisburd, 2016; Bernasco and Steenbeek, 2017) while retaining one of the most important and attractive properties — the interpretability — of Weisburd's original metric.

Using our proposed metric and data from New York City, Chicago and Philadelphia, three of the five largest cities in the United States, we show that while most types of crimes exhibit considerable concentration, the degree to which crimes are actually concentrated is smaller than has been suggested by prior analyses. For murder as well as common street crimes such as auto theft and robbery, we find that the law of crime concentration holds to a far lesser degree than has been supposed in prior literature which retains street segments that experienced at least one crime.

## 2 Prior Literature

### 2.1 Empirical Evidence on Crime Concentration

As noted by Weisburd (2015), the term "criminology of place" can be traced back to a 1989 article in *Criminology* by Sherman et al. (1989) which was among the first endeavors to systematically measure the concentration of crime among microgeographic areas.[5] However, the recognition that a large share of crime is clustered in a small share of places is an observation that is nearly as old as modern cities (Quetelet, 1831; Weisburd et al., 2009). Over the last few decades, a literature has proliferated to establish that micro- rather than macrogeography explains the lion's share of spatial variation in urban crime, (Steenbeek and Weisburd, 2016; Schnell et al., 2017), that crime is highly concentrated among a small number of crime hot spots (Eck et al., 2007; Weisburd, 2015) and that these hot spots, at least to an extent, persist over time (Weisburd et al., 2004, 2009; Gorr and Lee, 2015). Research has found that this pattern is not limited to low-impact crimes and applies equally, if not more forcefully, to some of the most costly criminal activity including gun crimes (Braga et al., 2010) and common street crimes such as robbery (Braga et al., 2010, 2011; Haberman et al., 2017)[6][7]

Since Weisburd's influential 2015 article, a rapidly growing literature, initiated by a 2017 special issue

[4]Randomizing crimes to blocks yields a result that is substantively similar to allocating crimes using a Poisson distribution.

[5]For a review of the historical development of the crime concentration literature, see Johnson (2010). For a review of theoretical work in this area, see Farrell (2015).

[6]Similarly, Hibdon et al. (2017) show that the law of crime concentration is substantively replicated when an additional data source — 911 calls for emergency service — are used to explore the concentration of illegal drug activity.

[7]These findings are subject to important criticisms regarding the measurement of crime concentration by Hipp and Kim (2017), Levin et al. (2017), Bernasco and Steenbeek (2017) and Curiel et al. (2018) which we discuss in Section 2.2.

on the criminology of place in the *Journal of Quantitative Criminology*, has developed to further test and clarify the law of crime concentration and the extent to which it holds across time and place. Recent scholarship documents robust evidence that the law of crime concentration substantively holds in other U.S. cities including Chicago (Schnell et al., 2017), Seattle (Hibdon et al., 2017), St. Louis (Levin et al., 2017) and a large number of cities in California (Hipp and Kim, 2017), in a number of non-U.S. cities including Vancouver, Canada (Andresen et al., 2017) Milan, Italy (Favarin, 2018) and among various cities in the United Kingdom (Oliveira et al., 2017) and Latin America (Ajzenman and Jaitman, 2016) as well as in a suburban setting — Brooklyn Park, Minnesota (Gill et al., 2017). In every setting in which the law of crime concentration has been tested, the law, as proposed, holds up substantively.

### 2.2 Conceptual Challenges to Measuring Crime Concentration

As noted by Andresen and Malleson (2011), Levin et al. (2017), Hipp and Kim (2017), Bernasco and Steenbeek (2017) Curiel et al. (2018) and Mohler et al. (2019) among others, the chief challenge to using the standard concentration metric is that it will lead to upward biased measures of crime concentration when the number of places is large relative to the number of crimes. That is, when crimes are relatively rare, it will be true by definition that a small number of places will account for most or even all of the crimes in a city.[8]

The literature has proposed two main ways of dealing with this issue. First, researchers have proposed shifting to a different metric — the Lorenz curve or the closely related Gini coefficient — that is more explicitly designed to measure the degree of inequality in a distribution (Bernasco and Steenbeek, 2017; OBrien, 2019; Mohler et al., 2019). A principal advantage of the Gini coefficient is that it allows researchers to characterize the relative degree of crime concentration using a single summary metric, without appealing to an arbitrary cutoff in the distribution of crimes (e.g., 25 or 50 percent of crimes). While the Gini coefficient can also perform poorly when crimes are sparse, recent research (Bernasco and Steenbeek, 2017; OBrien, 2019; Mohler et al., 2019) has proposed ways to address this concern, thus providing a means of measuring crime concentration that is potentially robust to the problem of crime-free street segments.[9] One limitation, however, in using

[8]This issue is not merely academic since, in many cities, a large number of street segments do not experience crime over a given time period (Curman et al., 2015) and, as it turns out, this issue has enormous implications for the conclusions that are drawn about which crimes are concentrated and the extent to which they are. For instance, as noted in Hipp and Kim (2017), while the standard crime concentration metric suggests that violent crimes are more concentrated than property crimes, after correcting the problem identified above, there is clear evidence that the degrees of concentration among violent and property crimes are, in fact, similar.

[9]Recent scholarship by Prieto Curiel (2019) raises questions about the stability of the Gini coefficient as well as the "modified Gini coefficient," noting that these are sensitive to changes in the underlying crime rate in a city.

the Gini coefficient is that it tends to be difficult to interpret, especially in communicating the degree to which criminal activity is concentrated to the wider world of criminal justice policymakers and researchers.[10] The Gini coefficient is, in our view, not very different from a correlation — it is an elegant and simple measure that is invariant to unit or scale — but it is difficult to communicate the cardinal information contained therein.

Given the inherent challenges in interpreting Gini coefficients, we argue that there remains a great deal of value in reporting crime concentration measures that correspond, as Weisburd proposed, to the share of places that account for a given share, typically one quarter or one half of crimes. Indeed, this remains a highly popular way to summarize crime concentration in the recent literature (Ajzenman and Jaitman, 2016; Gill et al., 2017; Andresen et al., 2017; Hibdon et al., 2017; Levin et al., 2017), sometimes alongside a Gini coefficient (Steenbeek and Weisburd, 2016; Schnell et al., 2017; Favarin, 2018; Vandeviver and Steenbeek, 2019; Umar et al., 2020).

Recent scholarship has proposed a second means of addressing issues caused by sparse crime data — while continuing to use Weisburd's original and highly interpretable crime concentration metric. The most popular solution in the literature which has been advanced in particular by Levin et al. (2017) is to measure the share of crimes that occur among the top $k$ percent of street segments, limiting the data to *the street segments that experienced at least one crime*. The intuition behind such a correction is straightforward: since many street segments do not experience any crime at all, these zero crime street segments will tend to make crime appear more concentrated than it actually is at the top of the distribution. Accordingly, the proposal is to focus on the segments in which crimes do occur. By addressing bias in measures of crime concentration that is an artifact of crime-free places, this proposed metric purports to move us closer to correctly estimating the extent to which crime is substantively concentrated. To see how this might work, consider a city in which half of street segments do not receive crime. If 2 percent of all street segments account for one quarter of the crimes, then it will be the case that 4 percent of street segments *which experience non-zero crime counts* account for one quarter of the crimes. Thus, the standard concentration metric will be two times too small.

### 2.3 Substantive Issues with Removing Crime-Free Segments

There is some virtue to removing crime free street segments — it is simple, easy to compute and understand and it does, in some applications, help to address the statistical artifact caused by sparse crime data.

[10]The Gini coefficient is formally 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.

However, as we demonstrate, this method will yield a metric of crime concentration that is biased upward – in many cases, considerably so. As we explain, the principal issue with this approach is that street segments which experience zero crimes is not the only reason why uniformity ($k$ percent of segments account for $k$ percent of crime) does not hold when crimes are assigned, at random, to street segments.

The implication of removing zero crime street segments to correct the non-uniformity problem is that, in the absence of any crime concentration, this measure of crime concentration should be 1. That is, the top $k$ percent of street segments that have non-zero crime should account for exactly $k$ percent of the crimes i.e., 25 percent of the street segments will account for 25 percent of the crimes, 50 percent of the street segments will account for 50 percent of the crimes, etc. This standard — that of uniformity — is an overly stringent standard that leads to an overestimate of the degree to which there is crime concentration. To see this, consider a simple example involving a city in which there are 1,000 street segments and 100 crimes. Using the standard measure of crime concentration, we would compute that $\frac{100}{1{,}000} = 10$ percent of street segments account for 100 percent of the crimes.

Using the metric advanced by Levin et al. (2017) and others, what would zero concentration look like? Zero concentration would hold if each crime occurred on a different street segment, as would be required under uniformity. However, a scheme in which crimes are randomized to street segments is unlikely — in fact, very unlikely — to produce the result that all 100 crimes occurred on different street segments (Curiel et al., 2018; Prieto Curiel, 2019). As a result, when this metric is applied to a dataset in which there is zero crime concentration by construction, it will indicate a positive amount of crime concentration. We show this using a simple simulation and later, in Section 5, we present evidence on the degree to which this metric yields an overestimate of crime concentration in empirical data.

## 3 An Adjusted Measure of Crime Concentration

In this section, we use randomization to propose a simple and easily interpretable way to identify the extent to which crimes are spatially concentrated.[11] We begin with a simple example and lay out our proposed framework. Consider a city that has $n$ street segments and experiences $j$ crimes. In the absence of concentration, what share of street segments *should* account for one quarter or one half of crimes? The rule

[11]Both Levin et al. (2017) and Hipp and Kim (2017) utilize simulation to elucidate the importance of a counterfactual in interpreting crime concentration statistics. However, neither paper utilizes randomization to generate a measure of marginal crime concentration. Recent work by Prieto Curiel (2019) shows, using simulated data, that probabilistic models — for instance the "rare event concentration coefficient" (RECC) — outperform the Gini coefficient in measuring crime concentration.

advanced by (Levin et al., 2017) suggests that we ought to expect uniformity — that is, $k$ percent of street segments account for $k$ percent of crimes. We consider the conditions under which this will be the case by running a simple simulation exercise. Consider a fictional city which has 1,000 street segments and a thought experiment in which the following number of crimes are assigned, at random with replacement, to these 1,000 street segments: 50, 100, 500, 1,000, 5,000, 10,000, 50,000, 100,000 and 1,000,000. What share of crimes would we expect to see represented among the top 25 percent of street segments, ranked according to the number of crimes experienced? Of course, under uniformity, we would expect that 25 percent and 50 percent of street segments to account for 25 percent and 50 percent of the crimes, respectively.

We present the results of the simulation exercise in **Figure 1**. In Figure 1, Panel A plots the share of all street segments and corresponds to the un-adjusted measure of crime concentration as was originally proposed by Weisburd (2015). Panel B plots the share of crime concentration among street segments that actually experience crime, as suggested by Levin et al. (2017) and others. In each panel, we plot the share of street segments accounting for 25 percent of the crimes using the dashed gray line and the share of street segments accounting for 50 percent of the crimes using the dashed black line. Horizontal reference lines are drawn at both 25 and 50 percent along the $y$-axis and represent the levels of crime concentration at which uniformity is achieved.

In Panel A, we see that when crime density is low relative to the number of street segments (e.g. $j = 50$ crimes amongst 1,000 segments), a very small share, approximately 1.2 percent of street segments, ranked by crime density, account for one quarter of the crimes. Likewise, just 2.4 percent of the street segments account for half of the crimes. As crimes become more common, each measure of crime concentration increases. When the number of crimes is 1,000 equaling the number of street segments, we see that 7.5 percent and 20 percent of street segments account for one quarter and one half of the crimes, respectively. At 10,000 crimes — or 10 crimes per street segment, approximately 15 percent of segments account for one quarter of the crimes and approximately 37 percent of segments account for one half of the crimes. At 1,000,000 crimes — 1,000 per street segment — uniformity is roughly met. As the number of crimes approaches infinity, uniformity will be achieved asymptotically. However, for relatively uncommon crimes or common crimes that are measured over a reasonably short window (e.g., one or two years), the asymptotic result will not hold and, as such, an un-adjusted measure of crime concentration will overstate the extent to which crimes are concentrated.

Next, we turn to Panel B which considers the performance of the popular metric which removes crime-free segments. At very low crime densities, this metric performs admirably. Conditioning on non-zero crime street segments leads to near-uniformity at 50 crimes for 1,000 street segments — here, 24.4 percent of the street segments account for one quarter of the crimes and 49 percent of the street segments account for 50 percent of the crimes. Likewise, this metric performs well asymptotically — though, of course, so does the un-adjusted metric. However, the metric performs far less well in the middle of the crime density distribution where we have between 1 and 100 crimes per street segment. For instance, at 1,000 crimes or 1 crime per street segment, we see that 12 percent of the segments account for one quarter of the crimes and 34 percent of the segments account for one half of the crimes. These figures are between one third and one half smaller than uniformity and the result is that crime concentration will be overestimated by between one third and one half. Likewise, at 10,000 crimes or 10 crimes per street segment, we see that 16 percent of segments account for one quarter of the crimes and approximately 39 percent of segments account for one half of the crimes. Incredibly, at these intermediate densities, removing crime-free street segments performs only marginally better than the un-adjusted metric. Since this window (between 1 and 10 crimes per street segment) is an extremely common density among the data that has been studied in the extant literature, the scope for the removal of crime-free segments to overstate crime density is unfortunately quite high.[12]

The thought experiment presented in Figure 1 makes clear that uniformity is an asymptotic result and does not hold in most empirical applications. We further see that removing the zero crime street segments does not substantively correct this issue at most crime densities. We thus propose a "corrected" metric of crime concentration that allows us to quantify the *marginal* degree of crime concentration above and beyond that which would be expected as an artifact of the density of the crime data:

$$mcc^k_{ijt} = cc^k_{ijt} - cc^{k*}_{ijt} \tag{1}$$

In (1), $mcc^k_{ijt}$ represents the marginal crime concentration in city $i$ for crime type $j$ over time period, $t$, and crime share $k$, where, for our purposes, $k =$ either 25 or 50 percent. $cc^{k*}_{ijt}$ is the crime concentration that is actually experienced in city $i$ (i.e., the measure proposed by Weisburd) for crime type $j$ and $cc^k_{ijt}$ is the crime concentration obtained under randomization with replacement. Since each randomized iteration of our randomization procedure will lead to a slightly different result, $cc^k_{ijt}$ will, in practice, be the mean crime concentration across a large number of trials. We later use variation across trials to generate

[12] In our data which spans between 10 and 15 years in three of the largest cities in the United States, the number of overall crimes per street segment varies between 35 and 115. Individual crime types are far less dense and vary between 0.5 and 10.

statistical inferences about crime concentration.

The larger is the value of $mcc_{ijt}^{k}$, the greater the degree of true crime concentration. Consider, for instance, a crime type for which $cc_{ijt}^{25} = 10$ percent and $cc_{ijt}^{25*} = 4$ percent. What this means is that, under the randomization of crimes to street segments, we would expect 10 percent of street segments to account for one quarter of the crimes. In reality, only 4 percent of street segments accounted for one quarter of the crimes. Hence, $mcc_{ijt}^{25} = 10$ percent - 4 percent = 6 percent. Accordingly, the additional share of street segments needed to account for one quarter of the crimes under randomization is 6 percent. Another way to express this is that crime is 2.5 times more concentrated than under randomization. Critically, unlike the standard crime concentration metric, higher *marginal* crime concentration indicates that crime is more concentrated. In Section 5, we estimate $mcc_{ijt}^{25}$ and $mcc_{ijt}^{50}$ for a variety of different crime types for each of our three cities: New York City, Chicago and Philadelphia.

## 4 Data

We derive estimates of the degree of marginal crime concentration using public crime microdata from three of the five largest cities in the United States: New York City (January 1st 2006 - December 31st, 2018), Chicago (January 1st, 2001 - May 4th, 2019) and Philadelphia (January 1st, 2006 - May 11th, 2019).[13, 14] We focus on a relatively long time period in order to better capture the spatial dynamics of crime. We note that, by focusing on a shorter time window — e.g., one year — the problem of non-uniformity will be even larger.

The data correspond to all crimes known to the city's municipal law enforcement agency. Each data set contains the coordinates where each crime occurred, allowing us to determine the street on which the crime happened.[15] The data also provide details on the type of offense, which we use to examine five categories on crime in addition to total crimes: murder, robbery, assault (simple and aggravated), motor vehicle theft, and larceny/theft. In keeping with prior literature, we assign crimes to street segments by determining which street is closest to the crime's coordinates through the mapping software ArcMap 10.6.1

[13]The crime datasets were downloaded from each city's Open Data website. Chicago: https://data.cityofchicago.org/Transportation/Street-Center-Lines/6imu-meau. New York City: https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i/data, https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Current-Year-To-Date-/5uac-w243/data. Philadelphia: https://www.opendataphilly.org/dataset/crime-incidents.

[14]We focus on these three cities because crimes from a fourth large city — Los Angeles — are coded primarily to intersections rather than street segments. Likewise, the city of Houston does not provide a shapefile of the city's street segments, excluding that city from the analysis.

[15]Fewer than 1 percent of crimes in each city have missing coordinates. As these crimes could not be matched to a street, they were removed from the data.

(Ratcliffe, 2012; Lee et al., 2017).[16,17] Following Weisburd (2015), we drop any crime that occurs in an intersection (i.e. matches with two or more street segments) or does not match to any street segments. We determine if a crime is near multiple streets by drawing 50-foot buffer around each street segment in the city and seeing how many streets are within 50-feet of each crime. There are substantial differences in the number of crimes geocoded to a single street segment rather than an intersection between each city. For both New York City (71 percent) and Chicago (94 percent), the majority of crime incidents are located within 50 feet of only one street segment, significantly larger than Philadelphia's 41 percent due to Philadelphia crimes being more commonly geocoded to a street intersection.

Out of concern that some street segments — such as highways — may not have any crimes coded to them or that extremely long street segments may have more crimes simply due to their length, we exclude any street that is 500 meters (1,640 feet) or longer. This excludes approximately 0.7 percent of streets in New York City, 0.3 percent of streets in Chicago, and 1.1 percent of streets in Philadelphia.[18] Because there is some uncertainty about whether some of the remaining street segments are, due to features of the physical environment, essentially "crime proof," we also engage in an auxiliary analysis in which we randomly remove 5 percent of crime-free street segments in the data are re-compute our measure of crime concentration. Results are substantively very similar.

We continue our discussion of the data by presenting descriptive statistics on crime in our three cities. **Table 1** presents, for each of our cities, the number of street segments as well as the number of crimes in the complete data set and in 2018, the last full year of data available. The cities included in this study vary widely with respect to the number of street segments in the city, though they are in rough accordance with the city's population. Philadelphia has slightly over 40,000 street segments, Chicago has about 56,000, and New York City has nearly 120,000. Chicago contains the largest number of crimes in our data, approximately 6.4 million, a function of the high crime rate in the city, the near complete matching of crime

[16]The street segment shapefiles were downloaded from each city's Open Data website. While some prior literature edits the shapefiles they used, we did not edit these files and used them exactly as downloaded from each city's website. Chicago: https://data.cityofchicago.org/Transportation/Street-Center-Lines/6imu-meau. New York City: https://data.cityofnewyork.us/City-Government/NYC-Street-Centerline-CSCL-/exjm-f27b. Philadelphia: https://www.opendataphilly.org/dataset/street-centerlines.

[17]Prior to merging each city's coordinates with the street segment shapefile, the crime data was projected to the proper coordinate reference system (CRS) based on the CRS of the given shapefile. To check the accuracy of both the coordinates and the merging process, a small number of coordinates were manually checked to ensure that they were located on the street they were merged to and that the address corresponded with the coordinates provided.

[18]Results are not sensitive to using a less conservative choice.

to a single street segment, and the fact that the available data extends as far back as 2001. Philadelphia and New York City have fewer crimes with 1 million and 4.6 million total crimes, respectively. While the total number of crimes differ between cities, the makeup of each city's crime is similar. Larceny in the most common crime in each city, consisting of between 21% (Chicago) and 27% (New York City) of crimes. In each city, murder is rare relative to other crimes, comprising just 0.1% of crimes reported in the city. These trends are roughly similar when examining crime that occurred in 2018, the last full year of data available.

Next, we consider crime concentration in each of our three cities, replicating the canonical figure from Weisburd (2015) which presents $cc_{ij}^{25*}$ and $cc_{ij}^{50*}$ for each of five large cities: Cincinnati OH, New York, NY, Sacramento, CA, Seattle, WA and Tel Aviv-Yafo (Israel). These data are presented in **Figure 2**, Panel A (crime concentration = 25 percent) and B (crime concentration = 50 percent). The gray bars represent the original cities in Weisburd's convenience sample. The black bars represent the three large cities for which we have data. Note that NYC is in both samples — the estimates differ slightly insofar as the sample years are slightly different. In Weisburd's convenience sample, in general, between 1-2 percent of street segments account for 25 percent of the crimes and between 4 and 6 percent of the street segments account for 50 percent of the crime, depending on the city. In our very large cities, crime is a little bit less concentrated but not dramatically so.

Crime is most concentrated in NYC which is relatively safe — 1.2 percent of street segments account for one quarter of the crimes and 4.6 percent of street segments account for one half of the crimes. Crimes are less concentrated in Chicago and Philadelphia which have higher levels of crime. In Chicago 2.8 percent of segments account for one quarter of the crimes and 9.3 percent of the segments account for one half of the crimes. In Philadelphia, those numbers are 2.2 percent and 8.6 percent respectively. Hence, the empirical regularity documented in Weisburd (2015) appears to roughly hold in our sample of cities too. In the next section, we characterize the extent to which crimes are concentrated, relative to what we argue is the ideal counterfactual — that which is generated using randomization with replacement rather than uniformity.

## 5 Results

### 5.1 Main Results

Our main results are presented in **Table 2A** and **Table 2B** which correspond with the share of street segments that account for one quarter and one half of crimes, respectively. The tables have a parallel structure and report key crime concentration metrics for each five crime types (murder, robbery, assault, motor vehicle theft and larceny) and aggregate crime in each of our three cities. Each table has five columns. The first column reports the proportion of street segments, when ranked in descending order of crime incidence, that account for $k$ percent of each type of crime. This is the standard (un-adjusted) measure of crime concentration referenced by Weisburd (2015). The second column reports the same quantity, conditioning on non-zero crime segments as is presented in the majority of the recent literature. The third column reports the same quantity in simulated data in which crimes are randomized to street segments, with replacement. The final two columns use the information in columns (1)-(3) to compute *marginal* crime concentration. Column (4) reports the measure of marginal concentration that we lay out in Section 3, equation (1). Column (5) reports the measure of crime concentration that is implied by the approach of removing crime-free street segments suggested by Levin et al. (2017).

We begin discussion of our findings by comparing the standard (un-adjusted) measure of crime concentration with crime concentration under randomization in each of our three cities. Consistent with computations presented in Weisburd (2015), in NYC, just over 1 percent of street segments account for one quarter of the crimes and just under 5 percent of street segments account for half of the crimes. In Chicago and Philadelphia, the figures are slightly higher but still imply a very large degree of crime concentration. For murder, the rarest crime in the data, the shares are especially small — across our three cities, between 0.3 and 1.1 of street segments account for one quarter of the murders and between 1-3 percent of the street segments account for half of the murders.

Next, we consider the simulated data in which crimes are randomly allocated to street segments. Here, we see that, for overall crimes, one quarter of the crimes would accrue to the top 19-22 percent of street segments and one half of the crimes would accrue to the top 42-46 percent of street segments. Given that we are using between 13 and 19 years of data in three of the largest cities in the United States, the data are sufficiently dense such that the distortionary impact of using the un-adjusted crime concentration metric to study total crime is fairly modest. For a smaller city or using a shorter time window, using the un-adjusted metric would result in far greater bias. However, for individual crime types, even when we use nearly two decades of data, the bias is large. For murder, in the simulated data, we observe that just 1-2 percent of street segments would account for one quarter of the murders under randomization. As such, the large degree of apparent crime concentration in the un-adjusted crime concentration metric is an artifact of the low density of homicides. The story is similar for auto thefts (25 percent of crimes accrue to 6.7 percent of street segments) and robberies (25 percent of crimes accrue to 8.1 percent of street segments) as well as for assaults and larcenies, albeit to a lesser degree. In each case, randomizing crimes to street segments artificially produces a pattern in the data that is consistent with crime concentration.

Next, we consider the method of removing crime-free street segments which has been suggested by Levin et al. (2017) and which has become a mainstay of the empirical literature. Conditioning on segments which experienced at least one crime, one quarter of crimes accrued to the top 2.2-3.4 percent of street segments and one half of the crimes accrued to the top 8.5-12 percent of street segments, depending on the city. Next, consider murder. Depending on the city, among street segments which experienced at least one murder, between 10-13 percent account for one quarter of murders implying that murder is concentrated to a considerable degree. However, we know that the empirical and simulated share of street segments which account for one quarter of crimes are nearly identical. As such, this simple comparison suggests that the method of removing crime-free segments does little to correct for the bias caused by low-density crime data.

What does this imply for our measure of marginal crime concentration and for the metric advanced by Levin et al. (2017) that excludes crime-free segments? Recall that, in Section 2.3, we claimed that the method of removing zero crime segments would yield a measure of crime concentration that is biased upward, a prediction that is supported by the simulations summarized in Figure 1. Here, we present empirical evidence for this claim using data from our three cities. Turning to columns (4) and (5) of Tables 2A and 2B, we compare our proposed metric — marginal crime concentration — to an equivalent metric which is implied by the method of removing crime-free street segments. Our metric (MCC) is reported in column (4) and is obtained by subtracting column (1) from column (3). The implied marginal crime concentration metric of Levin et al. (2017) is presented in column (5). Given that the metric implicitly assumes a counterfactual of uniformity ($k$ percent of street segments account for $k$ percent of crimes), marginal crime concentration can be computed by subtracting column (2) from either 25 percent (Table 2A) or 50 percent (Table 2B).

Using MCC, we see little evidence of crime concentration for murder and modest levels of crime concentration for robbery and auto theft. However, the method of Levin et al. (2017) implies considerable concentration for each of these crimes. For murder, the implied measure of crime concentration (between 11.6-15.1 depending on the city) is approximately 20 times larger than the measure of crime concentration which uses randomization to identify the counterfactual expectation (between 0.5-1.7 depending on the city). A similar story holds for auto theft where, for NYC, 7.6 percent of street segments which experienced auto theft account for one quarter of the auto thefts. These numbers imply that auto thefts are concentrated to a substantial degree. The marginal crime concentration metric implied by Levin et al. (2017) is 17.4. However, once we correctly account for the simulated distribution of crimes under randomization, we fail to see evidence of appreciable concentration — in NYC, auto thefts are concentrated by only four percentage points more than what would occur via randomization. Put differently, the method of crime-free street segments overstates the degree of crime concentration by between 80 and 300 percent, depending on the city. The method of Levin et al. (2017) likewise overstates the degree to which robberies are concentrated approximately 200 percent.

Turning to assault and larceny, we see considerable evidence of crime concentration albeit less than has been measured in the prior literature. The measures are quite similar by city, especially for overall crime, thus providing support for the idea that crime concentration may well be highly stable across cities. Overall, the method of removing crime-free segments yields as estimate of marginal crime concentration that is 50 percent too large for larceny and 67 percent too large for assaults.[19]

### 5.2 The Cumulative Density of Marginal Crime Concentration

A core virtue of Weisburd's approach to measuring crime concentration its simplicity and transparency. However, beyond the measurement issues which have been previously noted, an important limitation of focusing solely on the share of street segments which account for one quarter and one half of crimes is that such an approach is uninformative about the full distribution of crime concentration across the entire model space (Steenbeek and Weisburd, 2016). This limitation has given rise to a recent and growing literature which uses the Lorenz curve as a means of documenting the degree to which crimes are concentrated (Bernasco and Steenbeek, 2017; OBrien, 2019; Mohler et al., 2019). By plotting the cumulative density function of crimes against a line of perfect equality, the Lorenz curve allows researchers to generate a more complete understanding of the extent to which crimes are concentrated at each point along the distribution. At the same time, the key disadvantage of the Lorenz curve approach is that the associated Gini coefficient — which is the measure of crime concentration under the Lorenz curve — is extraordinarily difficult to interpret.

Recognizing that each of the two approaches brings something different to the table, in this section we present a graphical representation of marginal crime concentration by plotting the empirical and the simulated cumulative density function of crime together in the same figure. We present this graphical

[19]Given that there is some uncertainty about the presence of "crime proof" street segments in the data, in an auxiliary analysis presented in **Appendix Table 1A** and **Appendix Table 1B**, we have removed 5 percent of the crime-free street segments at random and re-computed our measure of marginal crime concentration. Results are extraordinarily similar to those reported in Tables 2A and 2B.

representation in **Figure 3A** (New York City), **Figure 3B** (Chicago) and **Figure 3C** (Philadelphia). In each figure, we plot the share of street segments ($y$-axis) that account for a given share of crimes ($x$-axis). The solid curve plots the empirical data. For an $x$-axis value of 100, the curve provides the share of crimes which account for all of the crimes of a given type in a given city. Consistent with Tables 2A and 2B, approximately 1-2 percent of street segments account for one quarter of the crimes and 5 percent of street segments account for half of the crimes. For rare crime categories like murder, the numbers are smaller. The dashed curve plots the null distribution given by share of street segments which account for a given share of crimes under randomization with replacement. We also provide a 45 degree line which represents perfect uniformity in which $k$ percent of street segments account for $k$ percent of crimes for all values of $k$.

The 45 degree line represents the implicit counterfactual under the method of removing crime-free street segments. The larger the vertical distance between the dashed simulated curve and the 45 degree line, the larger is the bias due to low density data. For a given value of $k$, marginal crime concentration is represented by the vertical distance between the dashed simulated curve and the solid empirical curve. Consistent with Tables 2A and 2B, for total crimes, there is a considerable degree of crime concentration which holds across the entire model space. For murder, crime concentration is extremely minimal with the solid and dashed lines lying almost directly on top of one another. For robbery and auto theft, crime concentration is likewise small, especially at the top of the distribution which, critically, is the range over which resources can potentially be deployed in a resource-constrained world.

In **Appendix Figure 1A**, **Appendix Figure 1B** and **Appendix Figure 1C**, we present the analog for the method of computing crime concentration by removing crime-free street segments which is a mainstay of the literature. Here, for a given value of $k$, the vertical distance between the 45 degree line and empirical crime concentration represents marginal crime concentration. Comparing Appendix Figures 2A-2C to Figures 3A-3C, it is easy to see that the method of removing crime-free segments appreciably overstates the degree to which there is crime concentration. This is particularly true for murder, robbery and auto theft. It is also notable that the degree to which the method of removing crime-free street segments leads to upwards bias in measured crime concentration is greatest at the top of the distribution which is precisely where this metric is most useful.

### 5.3 Statistical Inference

#### 5.3.1 Bootstrapped Confidence Interval

In computing crime concentration, a natural question to address is whether crime is *significantly* concentrated. That is, can we be confident that the empirical data are sufficiently different from the null distribution obtained via randomization? We compute bootstrapped 99 percent confidence intervals generated by re-randomizing crimes to street segments in 1,000 trials. For a given share of street segments, the boundaries of the confidence interval are formed by the 0.5th and 99.5th percentile of the bootstrapped distribution. For a given value of $k$, to the extent that the empirical share of street segments which account for $k$ percent of crimes lies outside the confidence interval, we can conclude that crime concentration is statistically significant.

In **Table 3**, we report the upper and lower limits of the 99 percent confidence interval along with the empirical share of street segments that account for 25 percent and 50 percent of crimes. For illustrative purposes, following Weisburd (2015), we focus on the share of street segments which account for 25 percent and 50 percent. However the substantive findings in Table 3 hold for each integer value of $k$ for all crime types. Two important lessons can be derived from the table. First, the confidence intervals are very narrow. Since we are using nearly two decades of data from three of the largest cities in the United States, the data are dense. Accordingly, there is very little variation in the spatial distribution of crimes that is achieved via random assignment. As such, we have extraordinary power in the data to test the null hypothesis of zero crime concentration. Second, for each crime type in each of our three cities, crime is significantly concentrated. Even murder — which is concentrated to only a very small degree substantively — is distributed significantly differently in practice than the null distribution implies.

#### 5.3.2 Pearson's $\chi^2$ Goodness-of-Fit Test

An alternative way to determine whether crimes are significantly concentrated is to directly compare the observed distribution of crimes in a city to the theoretical distribution arising from the random assignment of crimes to a city's street segments. The advantage of such an approach is that it can be used without needing to explicitly compute crime concentration. We do so using a version of Pearson's $\chi^2$ goodness-of-fit test which evaluates whether the distribution of the observed data differs significantly from a theoretical distribution (D'Agostino, 1986). In this case, the theoretical distribution of interest is a multinomial distribution which perfectly describes the problem given that we can think of each of our $k$ street segments as a category and each of the $n$ crimes as a trial, a scenario that is analogous to tossing a $k$-sided die $n$ times.

To generate a random allocation of crimes, we simulate data from random variables $X=(X_1,...,X_k)$, where $k$ is the number of possible mutually exclusive street segments with corresponding probabilities of a crime landing in a street segment, $p_1,...,p_k$. There are $n$ independent trials which represent the total number of crimes. In this case, $p_i=1/k$ since we assign equal probability of a crime occurring in any street segment. Since the $k$ outcomes are mutually exclusive, then $p_i \geq 0$ for $i=1,...,k$ and $\sum_{i=1}^{n} p_i=1$. Thus, we have that $X \sim \text{Multinom}(n,p_1,...,p_k)$.

The Pearson test statistic,$\chi^2$, is given by:

$$\chi^2=\sum_{i=1}^{n}\frac{(O_i-E_i)^2}{E_i}, \tag{2}$$

where $O_i$ is number of observed crimes per street segment $i$ where $i=1,...,k$, $E_i$ is the number of expected (or theoretical) crimes per street segment $i$, and $n$ is the total number of crimes. Given that the traditional $\chi^2$ test fails when there are many cells with small values, we simulate data from a multinomial distribution under the null hypothesis (that the observed distribution is multinomial), calculate the Pearson test statistic, and then repeat 1,000 times until we obtain the distribution of the test statistic. We then calculate the proportion of times that the test statistic from simulation is larger than the observed test statistic. This process yields an approximate $p$-value. Using this test, we find that in each of our three cities, there are statistically significant differences between the observed distribution of crimes and the random multinomial distribution ($p < 0.001$). We present results of this test for total crimes in **Appendix Figure 2**. However, consistent with our bootstrap inference test, this result also holds for each crime type that we study.

## 6 Conclusion

In this paper, we build upon recent methodological advances in the measurement of crime concentration and propose a method of measuring crime concentration that is simple, easy to interpret and is robust to a key statistical artifact — caused by sparse crime data — that the recent literature has been working to address. A common solution to this problem is to measure crime concentration *among street segments that actually experience crime*. This approach has become a mainstay of the recent literature (Ajzenman and Jaitman, 2016; Steenbeek and Weisburd, 2016; Gill et al., 2017; Andresen et al., 2017; Hibdon et al., 2017; Levin et al., 2017; Schnell et al., 2017; Favarin, 2018; Vandeviver and Steenbeek, 2019; Umar et al., 2020) and is often reported alongside a Gini coefficient. We note that while this approach will correct some of the upward bias in the measurement of crime concentration, appreciable bias will remain in most empirical applications. This is especially true at the top of the distribution of street segments. Therefore, the prior approach is most problematic for resource allocation problems where the constraints are the most binding.

Our proposed solution — comparing the actual distribution of crimes to a distribution of crimes under the randomization with replacement — allows us to generate a corrected measure of crime concentration that is robust to problems posed by sparse crime data. Our approach compliments recent advances which render Gini coefficient-based metrics robust to the same problem (Steenbeek and Weisburd, 2016; Bernasco and Steenbeek, 2017; OBrien, 2019; Mohler et al., 2019). However, a virtue of our approach is that it preserves the simplicity and interpretability of Weisburd's original crime concentration metric. While an advantage of the Gini coefficient relative to Weisburd's metric is that it allows researchers to characterize crime concentration without appealing to arbitrary cutoffs (e.g., 25 or 50 percent of crimes), our method easily lends itself to a full representation of the data. As such, we present a graphical representation of marginal crime concentration that offers a more interpretable analog to the Gini coefficient.

Our proposed metric — marginal crime concentration — is simply the excess share of crimes that occur in the top $k$ percent of locations relative to the null distribution. Using this metric, it is easy to compare observed crime concentration to random chance in an intuitive way — for instance, we might say that crime is 2-3 times more concentrated than it would be by chance. In this way, it is similar to positive predictive value (i.e., "precision"), a mainstay of prediction research in the social and computational sciences (James et al., 2013). Just as comparing the positive predictive value for a high-risk group to baseline risk in a population, we can compare crime concentration in actual versus simulated data in order to generate a metric that is simple and intuitive to understand.[20]

Like our predecessors, we find considerable evidence that crimes are concentrated among the cities we study and, accordingly, we provide additional support for the law of crime concentration. While the data suggests that each type of crime is *significantly* concentrated in a statistical sense, the extent to which the law of crime concentration applies requires qualification. In this research, we note that in three of the largest cities in the United States, while crime is highly concentrated in the aggregate, murders

[20] We further note that this methodology also has broad applicability to other domains in criminological research — for example to cohort studies of young people which invariably show that a small share of the population is responsible for an outsize share of the crimes and to "early warning systems" that police departments use to identify potentially problematic police officers and which are based on the premise that a small share of police officers are responsible for a disproportionate share of misconduct.

are effectively unconcentrated and the robberies and auto thefts are only concentrated to a very small degree. On the other hand, assaults and larcenies exhibit fairly substantial concentration at the city level. These results are qualitatively different than many of those in the literature and suggest that crime is less concentrated than has been suggested in the recent literature.

While marginal crime concentration solves a key problem that remains in the extant literature and while it is simpler and more easily interpreted than a Gini coefficient, several attendant limitations remain. First, like other measures of crime concentration, marginal crime concentration requires an assumption about the number of street segments in a city that are eligible to receive crimes. That is, some street segments that do not receive crimes over a given period may not simply be "lucky." It is possible that due to a natural feature of the built environment that some street segments are effectively "crime proof." To the extent that some street segments are immune to crime, our crime concentration metric has the potential to be biased. While we show empirically that the metric is effectively unchanged when we designate 5 percent of street segments to be ineligible to receive crimes, in applications involving a small number of street segments our metric may be more sensitive to such an assumption. As such, we recommend that researchers consider pre-processing their data in such applications. Second, while marginal crime concentration, like other approaches to crime concentration, allows an analyst to characterize the extent to which crime is concentrated in a given city over a given period of time, the policy implications that follow from a measure of crime concentration remain unclear. That is, the extent to which resources should be redeployed on the basis of spatial crime data depends not only on the extent to which crimes are concentrated among hot spots but also on the degree to which crimes are *ex ante* predictable and the extent to which crime is sensitive to the resources in question. While crime concentration remains a key estimand in allocating resources, it is best used alongside other information as well as domain-specific knowledge and common sense.

## References


Ajzenman, N. and L. Jaitman (2016). Crime concentration and hot spot dynamics in Latin America. Technical report, Inter-American Development Bank.

Andresen, M. A., A. S. Curman, and S. J. Linning (2017). The trajectories of crime at places: understanding the patterns of disaggregated crime types. *Journal of Quantitative Criminology 33*(3), 427–449.

Andresen, M. A. and N. Malleson (2011). Testing the stability of crime patterns: Implications for theory and policy. *Journal of Research in Crime and Delinquency 48*(1), 58–82.

Bernasco, W. and W. Steenbeek (2017). More places than crimes: Implications for evaluating the law of crime concentration at place. *Journal of Quantitative Criminology 33*(3), 451–467.

Bogar, S. and K. M. Beyer (2016). Green space, violence, and crime: A systematic review. *Trauma, Violence, & Abuse 17*(2), 160–171.

Bowers, K. (2014). Risky facilities: Crime radiators or crime absorbers? a comparison of internal and external levels of theft. *Journal of Quantitative Criminology 30*(3), 389–414.

Braga, A. A. (2001). The effects of hot spots policing on crime. *The ANNALS of the American Academy of Political and Social Science 578*(1), 104–125.

Braga, A. A. and B. J. Bond (2008). Policing crime and disorder hot spots: A randomized controlled trial. *Criminology 46*(3), 577–607.

Braga, A. A., D. M. Hureau, and A. V. Papachristos (2011). The relevance of micro places to citywide robbery trends: a longitudinal analysis of robbery incidents at street corners and block faces in boston. *Journal of Research in Crime and Delinquency 48*(1), 7–32.

Braga, A. A., A. V. Papachristos, and D. M. Hureau (2010). The concentration and stability of gun violence at micro places in boston, 1980–2008. *Journal of Quantitative Criminology 26*(1), 33–53.

Braga, A. A., A. V. Papachristos, and D. M. Hureau (2014). The effects of hot spots policing on crime: An updated systematic review and meta-analysis. *Justice Quarterly 31*(4), 633–663.

Braga, A. A., B. C. Welsh, and C. Schnell (2015). Can policing disorder reduce crime? a systematic review and meta-analysis. *Journal of Research in Crime and Delinquency 52*(4), 567–588.

Branas, C. C., R. A. Cheney, J. M. MacDonald, V. W. Tam, T. D. Jackson, and T. R. Ten Have (2011). A difference-in-differences analysis of health, safety, and greening vacant urban space. *American Journal of Epidemiology 174*(11), 1296–1306.

Branas, C. C., E. South, M. C. Kondo, B. C. Hohl, P. Bourgois, D. J. Wiebe, and J. M. MacDonald (2018). Citywide cluster randomized trial to restore blighted vacant land and its effects on violence, crime, and fear. *Proceedings of the National Academy of Sciences 115*(12), 2946–2951.

Chalfin, A., B. Hansen, J. Lerner, and L. Parker (2019). Reducing crime through environmental design: Evidence from a randomized experiment of street lighting in New York City. Technical report, National Bureau of Economic Research.

Chalfin, A., J. Kaplan, and M. LaForest (2020). Street light outages, public safety and crime displacement: Evidence from chicago. *Working Paper*.

Curiel, R. P., S. C. Delmar, and S. R. Bishop (2018). Measuring the distribution of crime and its concentration. *Journal of Quantitative Criminology 34*(3), 775–803.

Curman, A. S., M. A. Andresen, and P. J. Brantingham (2015). Crime and place: a longitudinal examination of street segment patterns in Vancouver, BC. *Journal of Quantitative Criminology 31*(1), 127–147.

D'Agostino, R. B. (1986). *Goodness-of-fit-techniques*, Volume 68. CRC press.

Davies, T. and S. D. Johnson (2015). Examining the relationship between road structure and burglary risk via quantitative network analysis. *Journal of Quantitative Criminology 31*(3), 481–507.

Doleac, J. L. and N. J. Sanders (2015). Under the cover of darkness: How ambient light influences criminal activity. *The Review of Economics and Statistics 97*(5), 1093–1103.

Eck, J. E., R. V. Clarke, R. T. Guerette, et al. (2007). Risky facilities: Crime concentration in homogeneous sets of establishments and facilities. *Crime Prevention Studies 21*, 225.

Farrell, G. (2015). Crime concentration theory. *Crime prevention and community Safety 17*(4), 233–248.

Farrington, D. P. and B. C. Welsh (2002). Improved street lighting and crime prevention. *Justice Quarterly 19*(2), 313–342.

Favarin, S. (2018). This must be the place (to commit a crime). testing the law of crime concentration in Milan, Italy. *European Journal of Criminology 15*(6), 702–729.

Garvin, E. C., C. C. Cannuscio, and C. C. Branas (2013). Greening vacant lots to reduce violent crime: a randomised controlled trial. *Injury Prevention 19*(3), 198–203.

Gill, C., A. Wooditch, and D. Weisburd (2017). Testing the law of crime concentration at place in a suburban setting: Implications for research and practice. *Journal of Quantitative Criminology 33*(3), 519–545.

Gorr, W. L. and Y. Lee (2015). Early warning system for temporary crime hot spots. *Journal of Quantitative Criminology 31*(1), 25–47.

Haberman, C. P., E. T. Sorg, and J. H. Ratcliffe (2017). Assessing the validity of the law of crime concentration across different temporal scales. *Journal of Quantitative Criminology 33*(3), 547–567.

Hibdon, J., C. W. Telep, and E. R. Groff (2017). The concentration and stability of drug activity in Seattle, Washington using police and emergency medical services data. *Journal of Quantitative Criminology 33*(3), 497–517.

Hipp, J. R. and Y.-A. Kim (2017). Measuring crime concentration across cities of varying sizes: Complications based on the spatial and temporal scale employed. *Journal of Quantitative Criminology 33*(3), 595–632.

James, G., D. Witten, T. Hastie, and R. Tibshirani (2013). *An introduction to statistical learning*, Volume 112. Springer.

Johnson, S. D. (2010). A brief history of the analysis of crime concentration. *European Journal of Applied Mathematics 21*(4-5), 349–370.

Keizer, K., S. Lindenberg, and L. Steg (2008). The spreading of disorder. *Science 322*(5908), 1681–1685.

Kelling, G. L., J. Q. Wilson, et al. (1982). Broken windows. *The Atlantic Monthly 249*(3), 29–38.

Kondo, M., B. Hohl, S. Han, and C. Branas (2016). Effects of greening and community reuse of vacant lots on crime. *Urban Studies 53*(15), 3279–3295.

Lee, Y., J. E. Eck, O. SooHyun, and N. N. Martinez (2017). How concentrated is crime at places? a systematic review from 1970 to 2015. *Crime Science 6*(1), 6.

Levin, A., R. Rosenfeld, and M. Deckard (2017). The law of crime concentration: An application and recommendations for future research. *Journal of Quantitative Criminology 33*(3), 635–647.

Mohler, G., P. J. Brantingham, J. Carter, and M. B. Short (2019). Reducing bias in estimates for the law of crime concentration. *Journal of Quantitative Criminology 35*(4), 747–765.

Moyer, R., J. M. MacDonald, G. Ridgeway, and C. C. Branas (2019). Effect of remediating blighted vacant land on shootings: a citywide cluster randomized trial. *American Journal of Public Health 109*(1), 140–144.

Oliveira, M., C. Bastos-Filho, and R. Menezes (2017). The scaling of crime concentration in cities. *PloS One 12*(8), e0183110.

OBrien, D. T. (2019). The action is everywhere, but greater at more localized spatial scales: Comparing concentrations of crime across addresses, streets, and neighborhoods. *Journal of Research in Crime and Delinquency 56*(3), 339–377.

Prieto Curiel, R. (2019). Is crime concentrated or are we simply using the wrong metrics? *arXiv preprint arXiv:1902.03105*.

Quetelet, A. (1831). Research on the propensity to crime of different ages brussels: Hayez.

Ratcliffe, J. H. (2012). The spatial extent of criminogenic places: a changepoint regression of violence around bars. *Geographical Analysis 44*(4), 302–320.

Schnell, C., A. A. Braga, and E. L. Piza (2017). The influence of community areas, neighborhood clusters, and street segments on the spatial variability of violent crime in Chicago. *Journal of Quantitative Criminology 33*(3), 469–496.

Sherman, L. W., P. R. Gartin, and M. E. Buerger (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology 27*(1), 27–56.

Sherman, L. W. and D. Weisburd (1995). General deterrent effects of police patrol in crime hot spots: A randomized, controlled trial. *Justice Quarterly 12*(4), 625–648.

Skogan, W. G. (2012). Disorder and crime. *The Oxford Handbook of Crime Prevention*, 173–188.

South, E. C., B. C. Hohl, M. C. Kondo, J. M. MacDonald, and C. C. Branas (2018). Effect of greening vacant land on mental health of community-dwelling adults: a cluster randomized trial. *JAMA Network Open 1*(3), e180298–e180298.

Steenbeek, W. and D. Weisburd (2016). Where the action is in crime? an examination of variability of crime across different spatial units in The Hague, 2001–2009. *Journal of Quantitative Criminology 32*(3), 449–469.

Umar, F., S. D. Johnson, and J. A. Cheshire (2020). Assessing the spatial concentration of urban crime: An insight from Nigeria. *Journal of Quantitative Criminology*, 1–20.

Vandeviver, C. and W. Steenbeek (2019). The (in)stability of residential burglary patterns on street segments: The case of Antwerp, Belgium 2005–2016. *Journal of Quantitative Criminology 35*(1), 111–133.

Weisburd, D. (2015). The law of crime concentration and the criminology of place. *Criminology 53*(2), 133–157.

Weisburd, D., G. J. Bruinsma, and W. Bernasco (2009). Units of analysis in geographic criminology: historical development, critical issues, and open questions. In *Putting Crime in its Place*, pp. 3–31. Springer.

Weisburd, D., S. Bushway, C. Lum, and S.-M. Yang (2004). Trajectories of crime at places: A longitudinal study of street segments in the city of Seattle. *Criminology 42*(2), 283–322.

Weisburd, D. and L. Green (1995). Policing drug hot spots: The Jersey City drug market analysis experiment. *Justice Quarterly 12*(4), 711–735.

Weisburd, D., N. A. Morris, and E. R. Groff (2009). Hot spots of juvenile crime: A longitudinal study of arrest incidents at street segments in Seattle, Washington. *Journal of Quantitative Criminology 25*(4), 443.

Weisburd, D. and C. W. Telep (2014). Hot spots policing: What we know and what we need to know. *Journal of Contemporary Criminal Justice 30*(2), 200–220.

Welsh, B. C. and D. P. Farrington (2008). Effects of improved street lighting on crime. *Campbell Systematic Reviews 13*, 1–51.

Table 1: Summary Statistics

| | New York City | Chicago | Philadelphia |
|---|---|---|---|
| Time period | Jan 2006 - Dec 2018 | Jan 2001 - May, 2019 | Jan 2006 - May, 2019 |
| Number of Street Segments | 118,653 | 56,179 | 40,542 |
| Percentage of Crime at Intersections | 29.0% | 5.6% | 59% |
| *Crime During Entire Studied Period* | | | |
| All crimes | 4,575,752 | 6,411,983 | 1,028,977 |
| Murder | 4,200 | 9,386 | 1,743 |
| Robbery | 131,179 | 237,369 | 28,669 |
| Assault | 683,463 | 1,591,870 | 219,45 |
| Auto theft | 81,903 | 296,415 | 20,451 |
| Larceny | 1,222,251 | 1,320,332 | 225,267 |
| *Crime During 2018* | | | |
| All crimes | 333,863 | 251,329 | 69,532 |
| Murder | 243 | 568 | 131 |
| Robbery | 8,394 | 9,158 | 1,665 |
| Assault | 55,566 | 67,320 | 16,247 |
| Auto theft | 3,925 | 9,556 | 1,034 |
| Larceny | 100,307 | 60,012 | 17,320 |

Note: Table presents descriptive
data on the number of street segments and crimes in each of our three cities: New York City, Chicago and Philadelphia.

Table 2A: Marginal Crime Concentration at 25 Percent of Street Segments

| | Share of Segments, Unadjusted (Weisburd) | Share of Segments, Non-Zero Crime Segments | Share of Segments, Simulated | Marginal Crime Concentration: Our proposed method | Marginal Crime Concentration: Levin-Rosenfeld-Deckard |
|---|---|---|---|---|---|
| A. New York City | | | | | |
| All crimes | 1.2 | 2.2 | 20.4 | 19.1 | 22.8 |
| Murder | .3 | 12.2 | .8 | .5 | 12.8 |
| Robbery | 1.1 | 4.3 | 8.1 | 7 | 20.7 |
| Assault | 1 | 2.5 | 14.8 | 13.8 | 22.5 |
| Auto theft | 2.2 | 7.6 | 6.7 | 4.5 | 17.4 |
| Larceny | .4 | .9 | 16.8 | 16.4 | 24.1 |
| B. Chicago | | | | | |
| All crimes | 2.8 | 3.4 | 22.2 | 19.4 | 21.6 |
| Murder | 1.1 | 9.9 | 2.9 | 1.7 | 15.1 |
| Robbery | 2.1 | 3.7 | 13.5 | 11.4 | 21.3 |
| Assault | 2.2 | 3.1 | 19.7 | 17.4 | 21.9 |
| Auto theft | 4.1 | 6.3 | 14.4 | 10.3 | 18.7 |
| Larceny | 1.3 | 1.7 | 19.2 | 17.9 | 23.3 |
| C. Philadelphia | | | | | |
| All crimes | 2.2 | 3.1 | 19.4 | 17.2 | 21.9 |
| Murder | .5 | 13.4 | 1 | .5 | 11.6 |
| Robbery | 1.1 | 4.1 | 6.8 | 5.6 | 21 |
| Assault | 2.1 | 4 | 14.6 | 12.5 | 21 |
| Auto theft | 2.1 | 8 | 5.5 | 3.4 | 17 |
| Larceny | .6 | 1 | 14.7 | 14.1 | 24 |

Note: This table reports the share of street segments that account for 25 and 50 percent of each of six crime types: total crime, murder, robbery, assault, auto theft and larceny and three cities: NYC (Panel A), Chicago (Panel B) and Philadelphia (Panel C). Column (1) reports crime concentration for all street segments, Column (2) reports crime concentration for street segments with non-zero crime and Column (3) reports simulated crime concentration arising generated by randomizing crimes to street segments, with replacement. The final two columns report *marginal* crime concentration using both our proposed method and by the method proposed by Levin-Rosenfeld-Deckard.

Table 2B: Marginal Crime Concentration at 50 Percent of Street Segments

| | Share of Segments, Unadjusted (Weisburd) | Share of Segments, Non-Zero Crime Segments | Share of Segments, Simulated | Marginal Crime Concentration: Our proposed method | Marginal Crime Concentration: Levin-Rosenfeld-Deckard |
|---|---|---|---|---|---|
| **A. New York City** | | | | | |
| All crimes | 4.6 | 8.5 | 43.6 | 39 | 41.5 |
| Murder | 1 | 36.3 | 1.7 | .7 | 13.7 |
| Robbery | 3.7 | 14.4 | 21 | 17.3 | 35.6 |
| Assault | 3.5 | 8.8 | 34.3 | 30.8 | 41.3 |
| Auto theft | 6.5 | 22.6 | 15.4 | 8.8 | 27.4 |
| Larceny | 2.9 | 5.9 | 38 | 35.1 | 44 |
| **B. Chicago** | | | | | |
| All crimes | 9.3 | 11.3 | 46.3 | 37 | 38.7 |
| Murder | 3.2 | 28.1 | 7 | 3.8 | 21.9 |
| Robbery | 7.4 | 13 | 32.3 | 24.9 | 37 |
| Assault | 7.2 | 10 | 42.6 | 35.4 | 40 |
| Auto theft | 11.8 | 18.2 | 33.9 | 22 | 31.8 |
| Larceny | 7.1 | 9.6 | 41.9 | 34.8 | 40.5 |
| **C. Philadelphia** | | | | | |
| All crimes | 8.6 | 11.9 | 42.2 | 33.6 | 38.1 |
| Murder | 1.5 | 40.9 | 2.1 | .6 | 9.1 |
| Robbery | 4.6 | 16.6 | 15.6 | 11 | 33.4 |
| Assault | 7 | 13.3 | 34 | 27 | 36.7 |
| Auto theft | 6.3 | 23.8 | 14.4 | 8.1 | 26.2 |
| Larceny | 5.5 | 9.3 | 34.1 | 28.5 | 40.8 |

Note: This table reports the share of street segments that account for 25 and 50 percent of each of six crime types: total crime, murder, robbery, assault, auto theft and larceny and three cities: NYC (Panel A), Chicago (Panel B) and Philadelphia (Panel C). Column (1) reports crime concentration for all street segments, Column (2) reports crime concentration for street segments with non-zero crime and Column (3) reports simulated crime concentration arising generated by randomizing crimes to street segments, with replacement. The final two columns report *marginal* crime concentration using both our proposed method and the method proposed by Levin-Rosenfeld-Deckard.

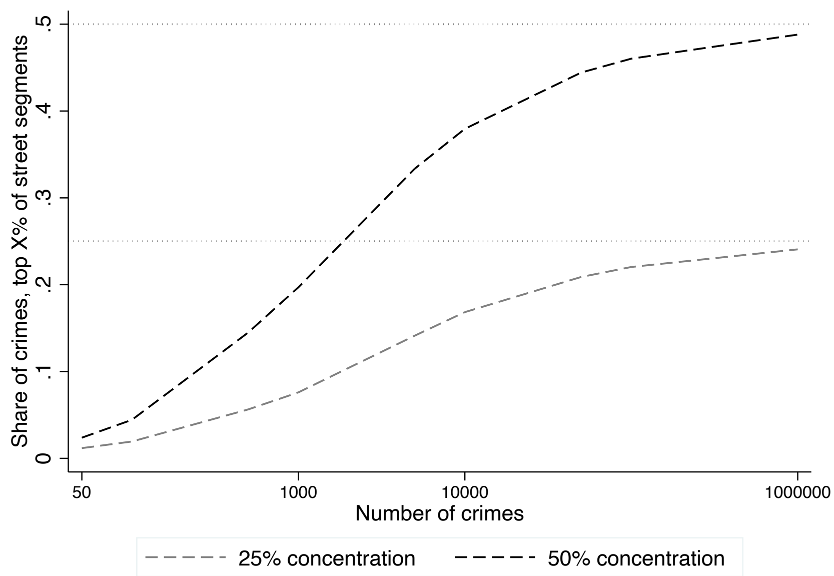
Table 3: Statistical Inference

| | Share of Crimes | Empirical Share of Street Segments | 99% Confidence Interval Lower Limit | 99% Confidence Interval Upper Limit |
|---|---|---|---|---|
| | | A. New York City | | |
| Total crime | 25 | .012 | .2032 | .2038 |
| | 50 | .0464 | .4359 | .4367 |
| Murder | 25 | .0034 | .008 | .0084 |
| | 50 | .0101 | .0169 | .0173 |
| Robbery | 25 | .0111 | .0807 | .0816 |
| | 50 | .0365 | .2085 | .2108 |
| Assault | 25 | .0097 | .1473 | .1486 |
| | 50 | .0347 | .3421 | .3437 |
| Auto theft | 25 | .0221 | .066 | .0673 |
| | 50 | .0653 | .1522 | .1554 |
| Theft | 25 | .0043 | .1677 | .1686 |
| | 50 | .0292 | .3798 | .3811 |
| | | B. Chicago | | |
| Total crime | 25 | .0277 | .2216 | .2221 |
| | 50 | .0933 | .4625 | .4631 |
| Murder | 25 | .0112 | .0275 | .0298 |
| | 50 | .0319 | .0692 | .0715 |
| Robbery | 25 | .0209 | .1341 | .1358 |
| | 50 | .074 | .3213 | .3241 |
| Assault | 25 | .0224 | .1963 | .1972 |
| | 50 | .0722 | .4256 | .4268 |
| Auto theft | 25 | .0411 | .1438 | .1452 |
| | 50 | .1182 | .3373 | .3399 |
| Theft | 25 | .0126 | .1916 | .1926 |
| | 50 | .0706 | .4183 | .4197 |
| | | C. Philadelphia | | |
| Total crime | 25 | .0223 | .1935 | .1945 |
| | 50 | .0855 | .4213 | .4229 |
| Murder | 25 | .0049 | .0094 | .0102 |
| | 50 | .0148 | .0202 | .021 |
| Robbery | 25 | .0112 | .066 | .0687 |
| | 50 | .046 | .1544 | .1576 |
| Assault | 25 | .0208 | .1448 | .1466 |
| | 50 | .0703 | .3384 | .3414 |
| Auto theft | 25 | .021 | .0537 | .0556 |
| | 50 | .0626 | .1409 | .1468 |
| Theft | 25 | .0063 | .1461 | .1477 |
| | 50 | .0553 | .3394 | .3422 |

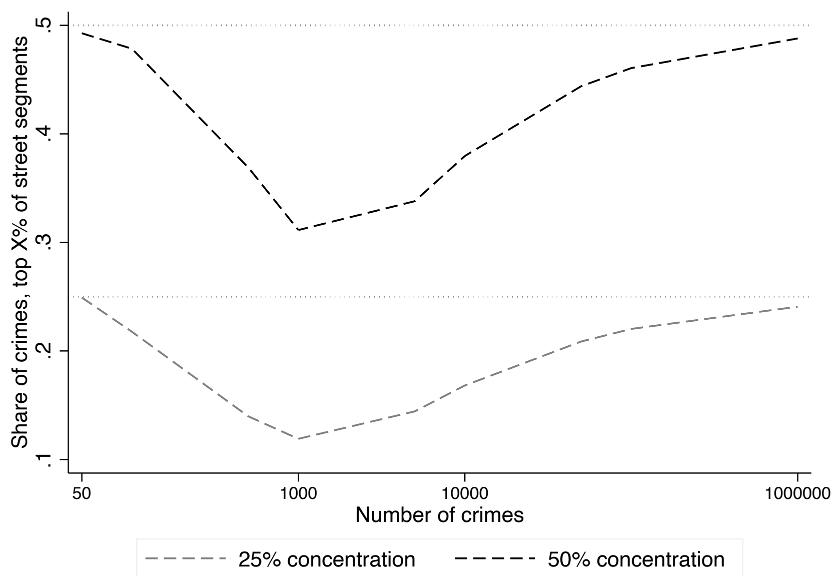
Note: For each crime type, in each city, this table reports the share of street segments that account for 25 percent and 50 percent of the crimes as well as the lower and upper limits of the 99 percent bootstrapped confidence interval calculated through a re-randomization routine in which crimes are randomly assigned to street segments, with replacement.

Figure 1: Crime Concentration, Simulated Data for $n$ = 1,000 street segments

Panel A: Unadjusted crime concentration, all street segments



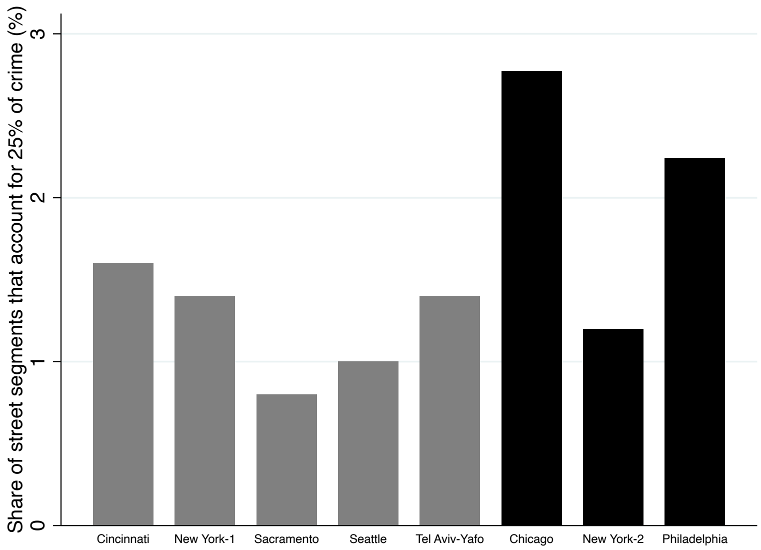

Panel B: Unadjusted crime concentration, non-zero crime street segments



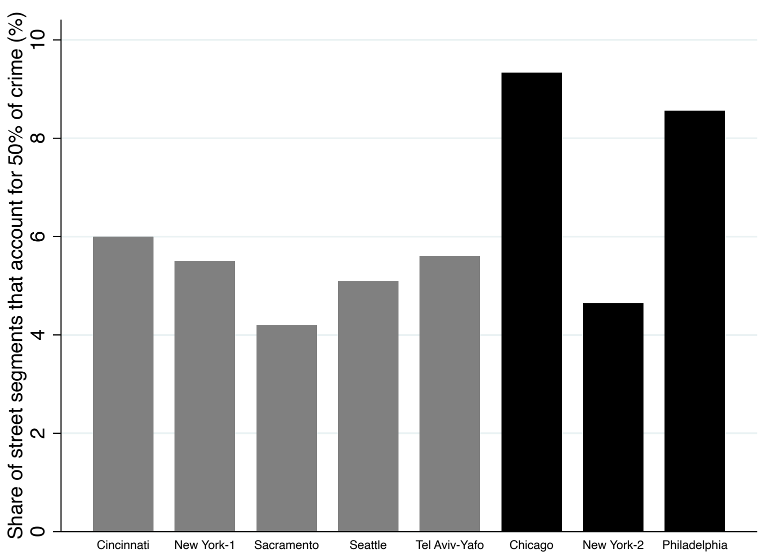

Note: Figures plot the share of street segments that account for 25 percent (Panel A) and 50 percent (Panel B) of crimes, in simulated data in which crimes are randomly assigned to street segments, with replacement. The number of street segments is fixed at 1,000 while the number of crimes is allowed to very along the $x$-axis. The $x$-axis has been transformed using a logarithmic scale.

Figure 2: Share of Crimes Among the Top 25 and 50 Percent of Street Segments, by City

Panel A: Unadjusted crime concentration, all street segments

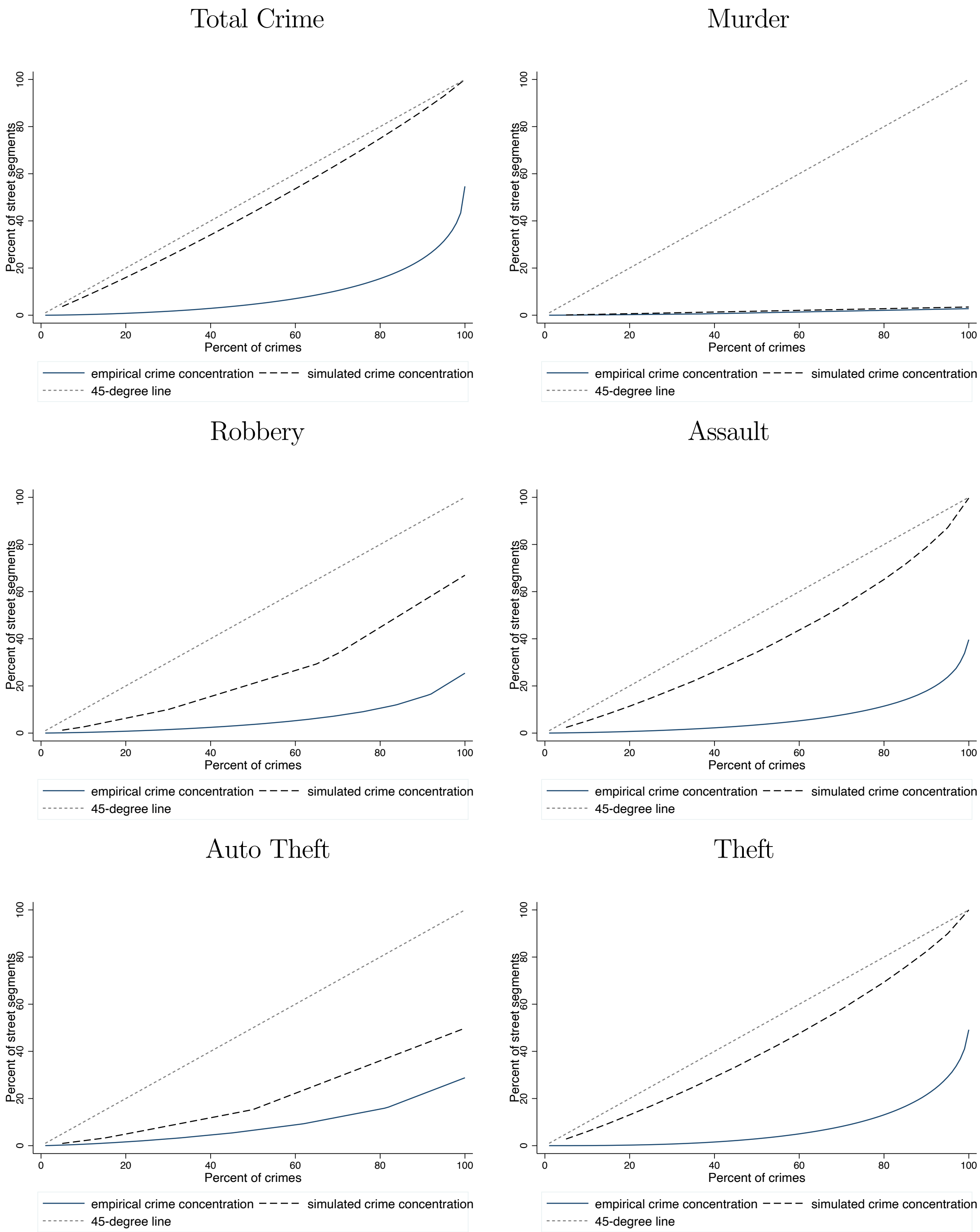



Panel B: Unadjusted crime concentration, non-zero crime street segments



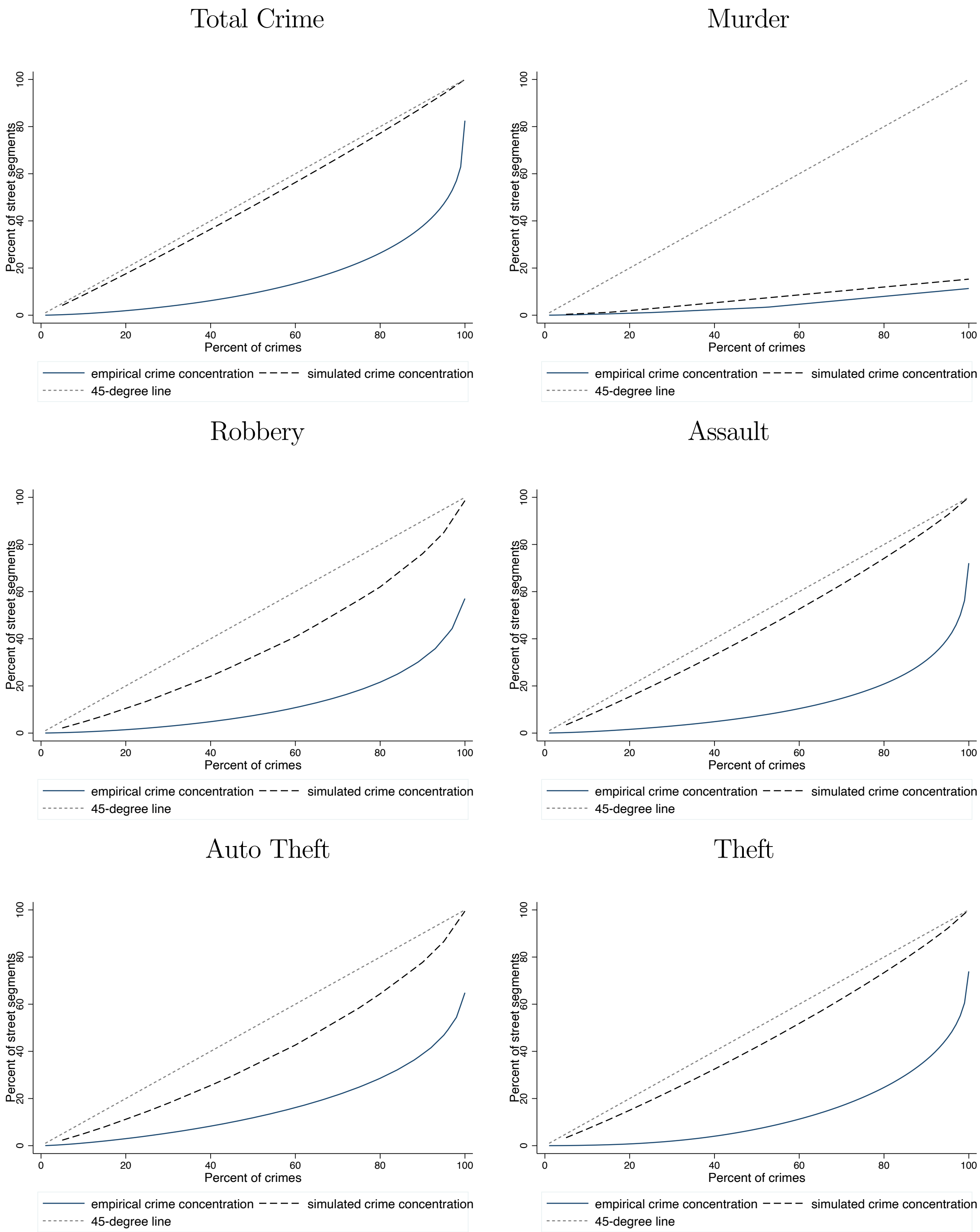

Note: Figures plot the share of
street segments that account for 25 percent (Panel A) and 50 percent (Panel B) of crimes. The gray bars are replicated from Table 3 in Weisburd (2015). The black bars correspond to data from New York City, Chicago and Philadelphia (our sample).

Figure 3A: Cumulative Density of Crime Concentration (New York City)



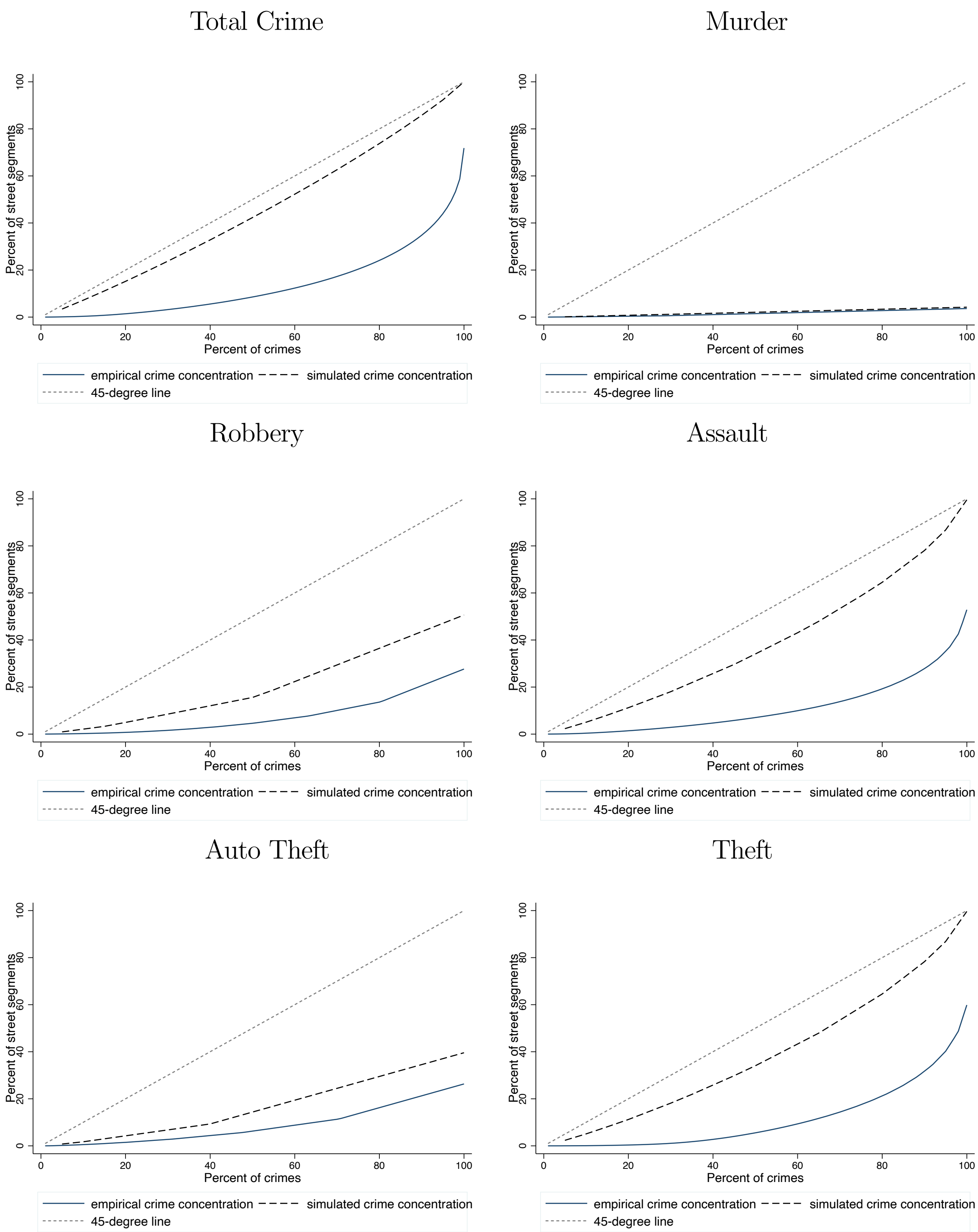

Note: Figures plot the share of street segments ($y$-axis) that account for a given percentage of crimes ($x$-axis). The solid curve represents the empirical data, the dashed curve represents simulated data in which crimes are randomized to street segments, with replacement. The gray dotted line is a 45 degree line. For a given value of $k$, crimes are concentrated when the heights of the solid and dashed curves are different. Marginal crime concentration is the vertical distance between the dashed and the solid curves.

Figure 3B: Cumulative Density of Crime Concentration (Chicago)




Note: Figures plot the share of street segments ($y$-axis) that account for a given percentage of crimes ($x$-axis). The solid curve represents the empirical data, the dashed curve represents simulated data in which crimes are randomized to street segments, with replacement. The gray dotted line is a 45 degree line. For a given value of $k$, crimes are concentrated when the heights of the solid and dashed curves are different. Marginal crime concentration is the vertical distance between the dashed and the solid curves.

Figure 3C: Cumulative Density of Crime Concentration (Philadelphia)




Note: Figures plot the share of street segments ($y$-axis) that account for a given percentage of crimes ($x$-axis). The solid curve represents the empirical data, the dashed curve represents simulated data in which crimes are randomized to street segments, with replacement. The gray dotted line is a 45 degree line. For a given value of $k$, crimes are concentrated when the heights of the solid and dashed curves are different. Marginal crime concentration is the vertical distance between the dashed and the solid curves.

# Online Appendix Material

Appendix Table 1A:
Marginal Crime Concentration at 25 Percent of Street Segments, Removing 5 Percent of Street Segments

| | | | | Marginal Crime Concentration | |
|---|---|---|---|---|---|
| | Share of Segments, Unadjusted (Weisburd) | Share of Segments, Non-Zero Crime Segments | Share of Segments, Simulated | Our proposed method | Levin-Rosenfeld-Deckard |
| **A. New York City** | | | | | |
| All crimes | 1.2 | 2.2 | 20.5 | 19.3 | 22.8 |
| Murder | .3 | 12.3 | .9 | .5 | 12.8 |
| Robbery | 1.1 | 4.4 | 8.4 | 7.3 | 20.6 |
| Assault | 1 | 2.5 | 14.9 | 14 | 22.5 |
| Auto theft | 2.2 | 7.7 | 6.8 | 4.6 | 17.3 |
| Larceny | .4 | .9 | 17 | 16.6 | 24.1 |
| **B. Chicago** | | | | | |
| All crimes | 2.8 | 3.4 | 22.3 | 19.5 | 21.6 |
| Murder | 1.1 | 9.9 | 2.9 | 1.8 | 15.1 |
| Robbery | 2.1 | 3.7 | 13.8 | 11.7 | 21.3 |
| Assault | 2.2 | 3.1 | 19.8 | 17.5 | 21.9 |
| Auto theft | 4.1 | 6.3 | 14.7 | 10.6 | 18.7 |
| Larceny | 1.3 | 1.7 | 19.4 | 18.1 | 23.3 |
| **C. Philadelphia** | | | | | |
| All crimes | 2.2 | 3.1 | 19.5 | 17.3 | 21.9 |
| Murder | .5 | 13.4 | 1 | .5 | 11.6 |
| Robbery | 1.1 | 4.1 | 6.9 | 5.8 | 21 |
| Assault | 2.1 | 4 | 14.8 | 12.7 | 21 |
| Auto theft | 2.1 | 8 | 5.7 | 3.6 | 17 |
| Larceny | .6 | 1 | 14.8 | 14.2 | 24 |

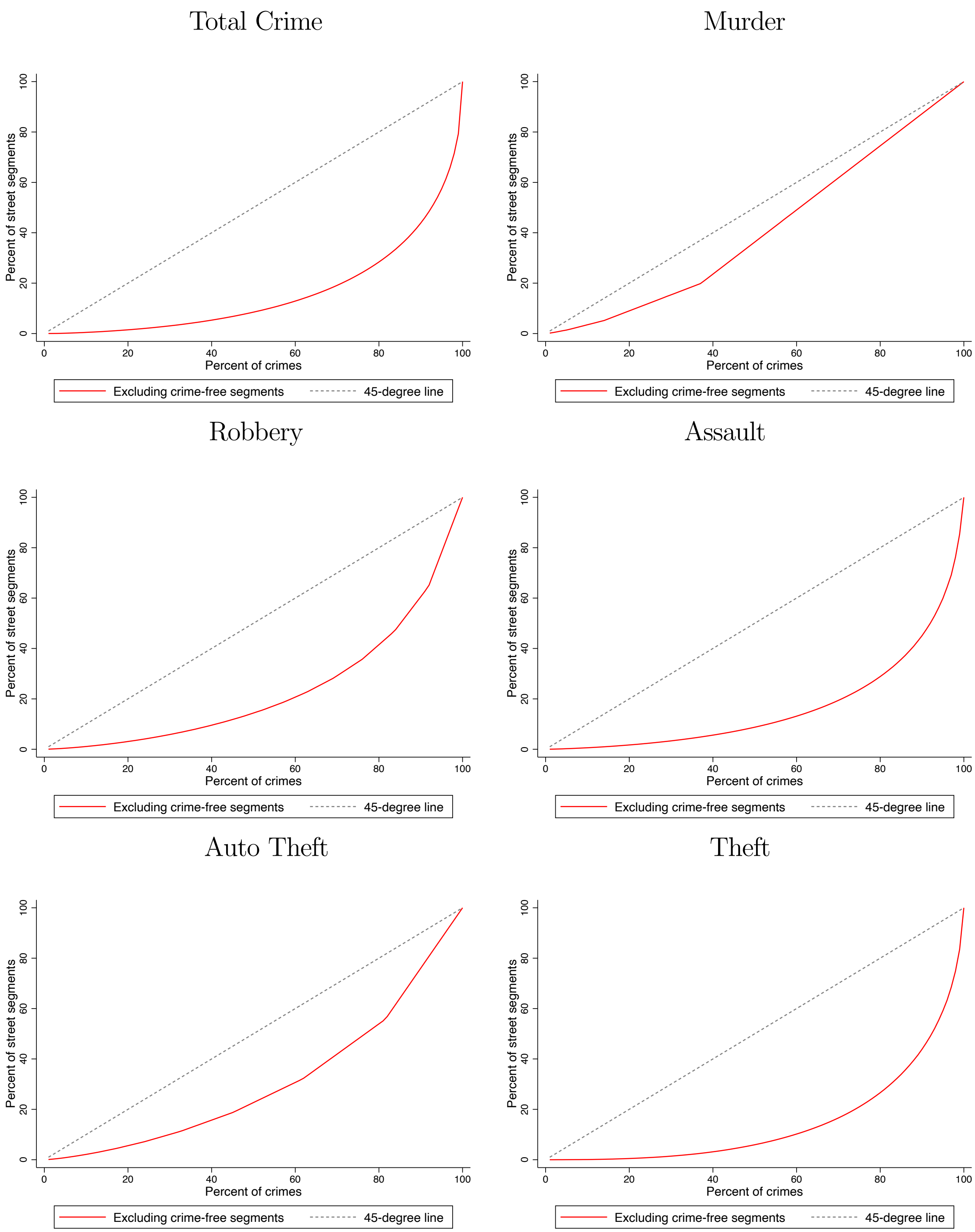
Note: This table reports the share of street segments that account for 25
and 50 percent of each of six crime types: total crime, murder, robbery, assault, auto theft and larceny and three cities: NYC (Panel A), Chicago (Panel B) and Philadelphia (Panel C). Column (1) reports crime concentration for all street segments, Column (2) reports crime concentration for street segments with non-zero crime and Column (3) reports simulated crime concentration arising generated by randomizing crimes to street segments, with replacement. The final two columns report *marginal* crime concentration using both our proposed method and by the method proposed by Levin-Rosenfeld-Deckard.

Appendix Table 1B:
Marginal Crime Concentration at 50 Percent of Street Segments, Removing 5 Percent of Street Segments

| | Share of Segments, Unadjusted (Weisburd) | Share of Segments, Non-Zero Crime Segments | Share of Segments, Simulated | Marginal Crime Concentration: Our proposed method | Marginal Crime Concentration: Levin-Rosenfeld-Deckard |
|---|---|---|---|---|---|
| **A. New York City** | | | | | |
| All crimes | 4.7 | 8.5 | 43.8 | 39.1 | 41.5 |
| Murder | 1 | 36.4 | 1.8 | .8 | 13.6 |
| Robbery | 3.7 | 14.4 | 21.5 | 17.9 | 35.6 |
| Assault | 3.5 | 8.8 | 34.8 | 31.3 | 41.3 |
| Auto theft | 6.5 | 22.7 | 15.9 | 9.4 | 27.3 |
| Larceny | 2.9 | 5.9 | 38.3 | 35.3 | 44 |
| **B. Chicago** | | | | | |
| All crimes | 9.3 | 11.3 | 46.4 | 37 | 38.7 |
| Murder | 3.2 | 28.1 | 7.3 | 4.1 | 21.9 |
| Robbery | 7.4 | 13 | 32.5 | 25.1 | 37 |
| Assault | 7.2 | 10 | 42.8 | 35.6 | 40 |
| Auto theft | 11.8 | 18.2 | 34.1 | 22.3 | 31.8 |
| Larceny | 7.1 | 9.6 | 42.1 | 35 | 40.5 |
| **C. Philadelphia** | | | | | |
| All crimes | 8.6 | 11.9 | 42.4 | 33.8 | 38.1 |
| Murder | 1.5 | 40.9 | 2.2 | .7 | 9.1 |
| Robbery | 4.6 | 16.6 | 16.2 | 11.6 | 33.4 |
| Assault | 7 | 13.3 | 34.2 | 27.1 | 36.7 |
| Auto theft | 6.3 | 23.8 | 14.6 | 8.4 | 26.2 |
| Larceny | 5.5 | 9.3 | 34.5 | 28.9 | 40.8 |

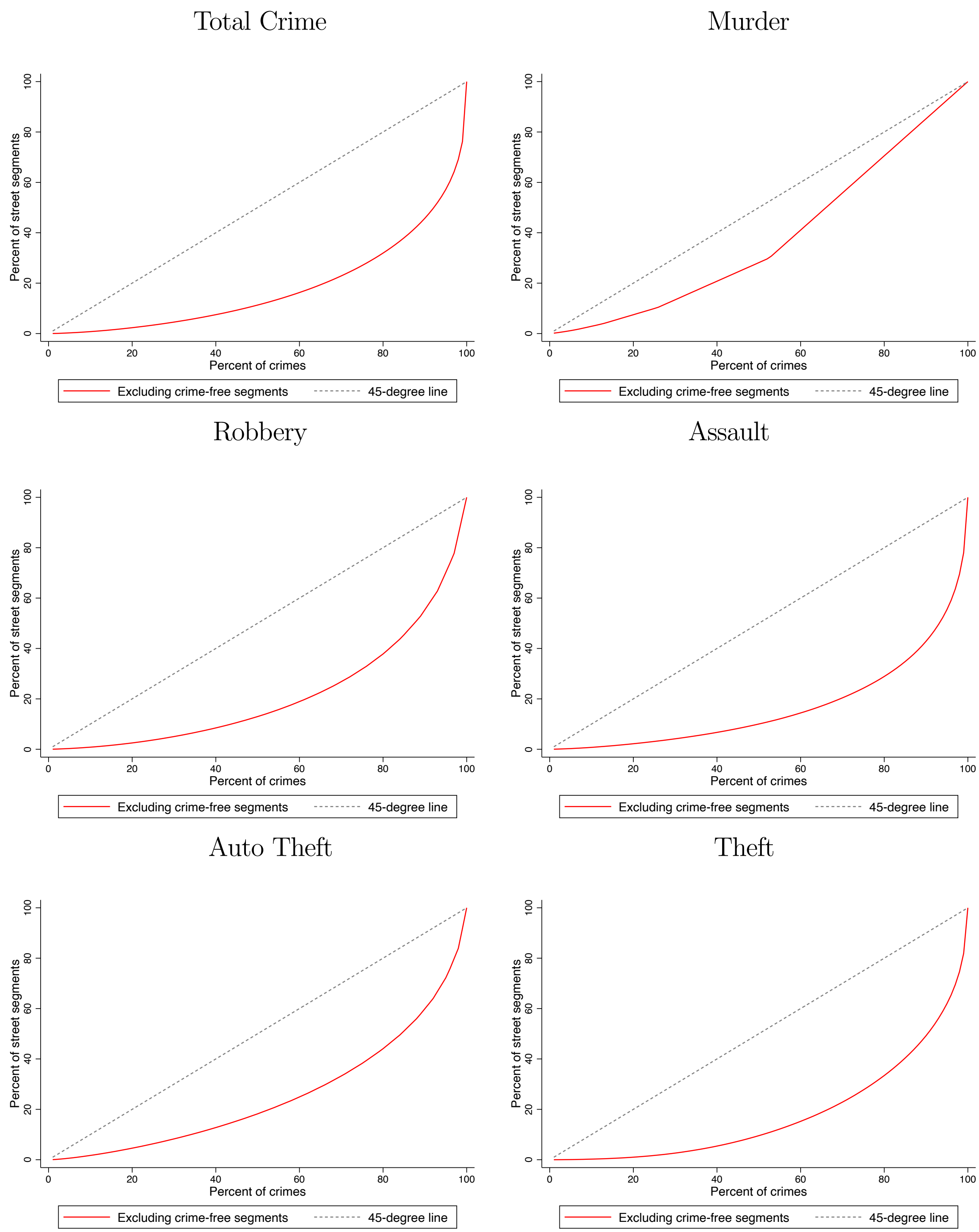
Note: This table reports the share of street segments that account for 25 and 50 percent of each of six crime types: total crime, murder, robbery, assault, auto theft and larceny and three cities: NYC (Panel A), Chicago (Panel B) and Philadelphia (Panel C). Column (1) reports crime concentration for all street segments, Column (2) reports crime concentration for street segments with non-zero crime and Column (3) reports simulated crime concentration arising generated by randomizing crimes to street segments, with replacement. The final two columns report *marginal* crime concentration using both our proposed method and the method proposed by Levin-Rosenfeld-Deckard.

Appendix Figure 1A: Cumulative Density of Crime Concentration, Removing Crime-Free Street Segments (New York City)



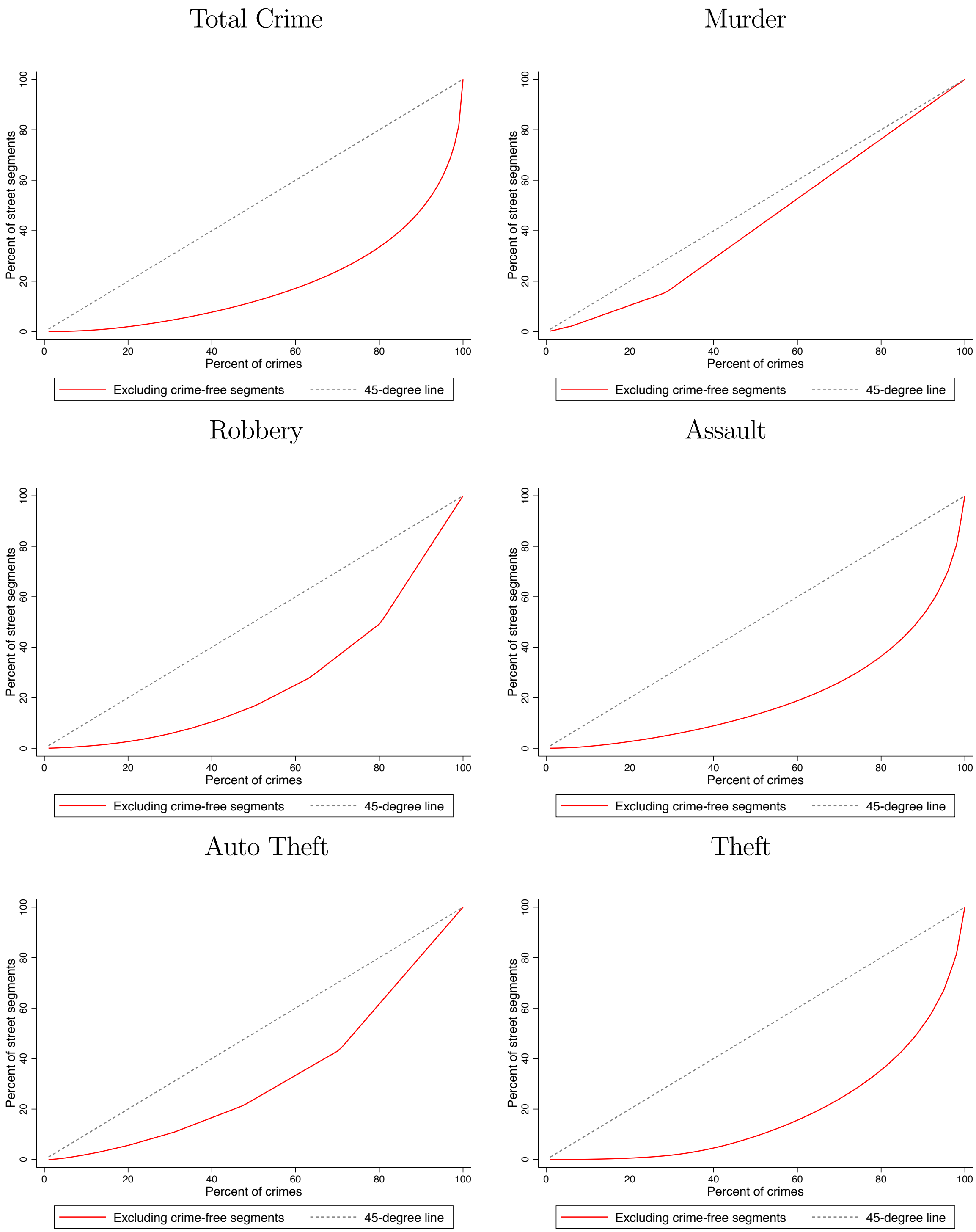

Note: Figures plot the share of street segments ($y$-axis) that account for a given percentage of crimes ($x$-axis). The solid red curve represents the empirical data, after removing crime-free street segments. The dotted line is a 45 degree line. Under the approach of Levin et al. (2017), for a given value of $k$, crimes are concentrated when the heights of the two curves are different.

Appendix Figure 1B: Cumulative Density of Crime Concentration,
Removing Crime-Free Street Segments (Chicago)



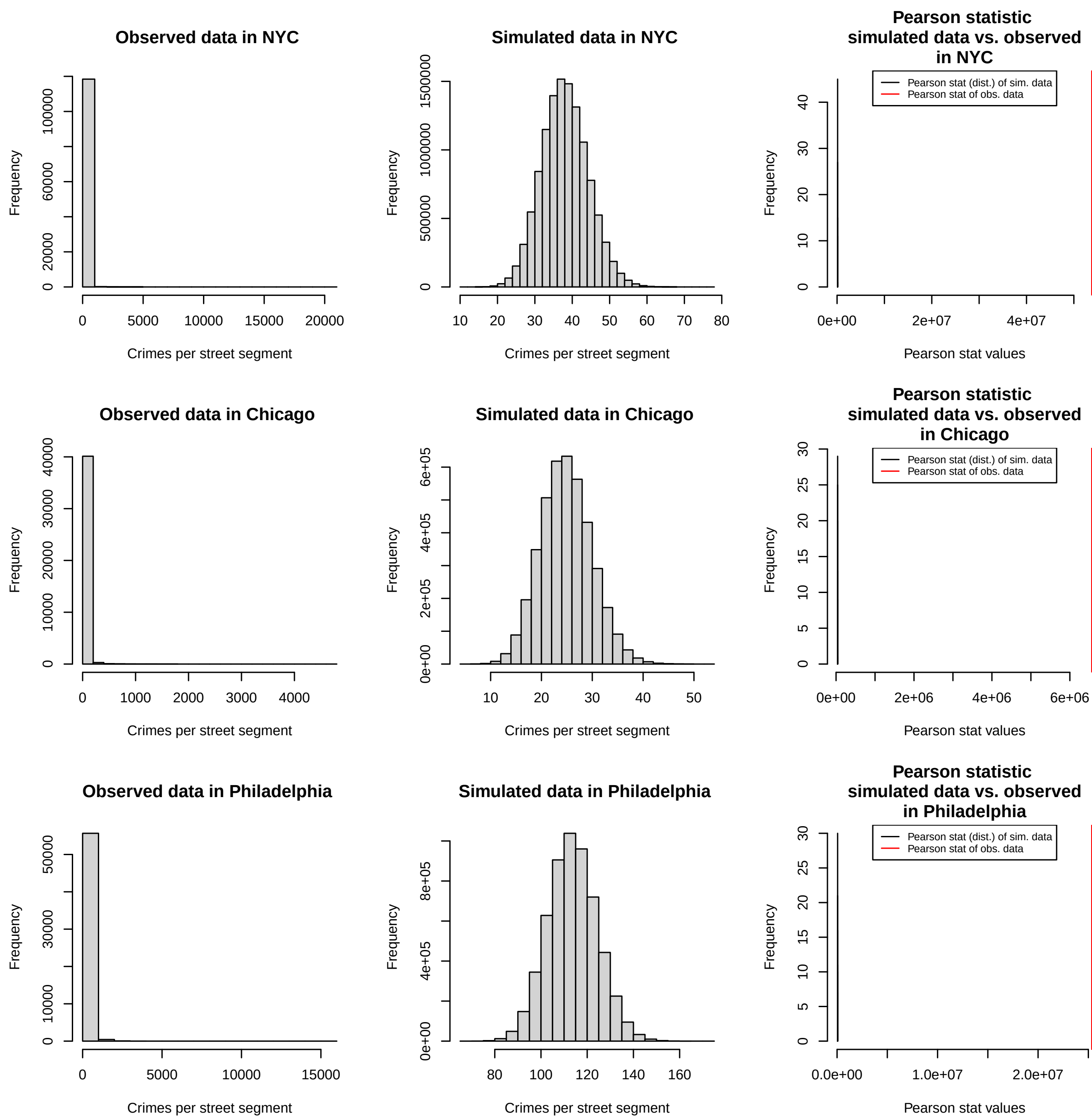

Note: Figures plot the share of street segments ($y$-axis) that account for a given percentage of crimes ($x$-axis). The solid red curve represents the empirical data, after removing crime-free street segments. The dotted line is a 45 degree line. Under the approach of Levin et al. (2017), for a given value of $k$, crimes are concentrated when the heights of the two curves are different.

Appendix Figure 1C: Cumulative Density of Crime Concentration, Removing Crime-Free Street Segments (Philadelphia)




Note: Figures plot the share of street segments ($y$-axis) that account for a given percentage of crimes ($x$-axis). The solid red curve represents the empirical data, after removing crime-free street segments. The dotted line is a 45 degree line. Under the approach of Levin et al. (2017), for a given value of $k$, crimes are concentrated when the heights of the two curves are different.

Appendix Figure 2: Pearson's $\chi^2$ Test: Observed versus Simulated Spatial Crime Distributions




Note: Figures plot the simulated (according to a multinomial distribution) and observed distributions of total crime for each of our three cities — NYC, Chicago and Philadelphia. We also plot the Pearson $\chi^2$ test statistic from the simulated and observed data. In all cases, the observed statistic lies far from the distribution of simulated Pearson statistics, and thus the observed data distribution is significantly different from the simulated data distribution.