Eyes on the Street, Spatial Concentration of Retail Activity and Crime

Stuart S. Rosenthal
Paul Rubacha Department of Real Estate
Cornell University, Ithaca, New York
Email: sr784@cornell.edu

Joaquin A. Urrego
Haas School of Business
University of California – Berkeley, Berkeley, CA
Email: jaurrego@berkeley.edu

December 1, 2023

* We thank Leah Brooks, Giulia Faggio, William Horrace, and Amy Schwartz for helpful comments, in addition to seminar participants at the University of Alabama, the 2021 North American UEA conference and the 2022 AREUEA/ASSA conference. We also thank Safegraph for access to cellphone data including Mike Sussman, Mitch McGillis, and Jonathan Wolf. All errors are our own.

Abstract

If spatial concentration of retail establishments amplifies the effect of "eyes on the street", that should lower neighborhood crime rates and reduce investment in anti-crime measures, with benefits capitalized into higher retail rent. Data for New York City supports these predictions. In addition, comparisons between nighttime versus daytime crime, pre-pandemic versus COVID-19 lockdown, and different measures of spatial concentration shed light on mechanisms. Under plausible identifying conditions, increasing neighborhood concentration of retail outlets by one standard deviation reduces property crime and police stops by at least 8.5% and 11%, respectively, and causes retail rent to increase by at least 7.8%.

JEL Codes: R00, R30, K00.
Key Words: Eyes-on-the-street, Spatial Concentration, Crime, Police, Retail, Rent Capitalization

1. Introduction

Longstanding arguments suggest that "eyes on the street" deter crime by increasing the likelihood that criminals will be caught (Jacobs, 1961; Browning and Jackson, 2013; Chang and Jacobson, 2017; Carr and Doleac, 2018; McMillen et al., 2019). An alternate view is that criminals are better able to hide in a crowd (Jarrell and Howsen, 1990; Harries, 2006; Tillyer and Walter, 2019). We consider these issues, focusing primarily on the effect of within-neighborhood spatial patterns of retail outlets on property crime.[1] Retail outlets have valuable inventory that attracts crime, but also draw crowds of shoppers. Based on a simple conceptual model, we argue that if concentrating retail establishments at the street level amplifies the effect of eyes on the street – a type of neighborhood-level external economy of scale – this should deter crime and reduce public and private investment in anti-crime measures, with benefits capitalized into higher retail rent.[2] Estimates based on point-specific data for New York City support these priors, an implication of which is that local government can reduce the cost of crime by encouraging spatial concentration of retail activity.[3]

The potential for spatial concentration of retail establishments to reduce the cost of crime is large. A 2018 National Retail Federation (NRF) survey of U.S. retailers found that respondents lost an average of 1.38% of sales to theft of merchandise and services, similar to a 1.3% loss rate reported for retailers in Europe in the 2018-2020 World Bank Enterprise survey.[4] These same surveys indicate that U.S. retailers allocated 0.74% of sales to private security measures while European retailers spent roughly 0.8% of sales. Benchmarked against 2018 retail sales in NYC (roughly $100 billion), the NRF estimates suggest

---

[1] Property crime includes petit larceny, grand larceny, burglary, theft of services and fraud. In some instances, we also consider auto theft and robbery which are classified separately.

[2] As described above, eyes on the street contribute to neighborhood level productivity spillovers but in ways that differ from mechanisms typically highlighted in the agglomeration literature (e.g., Rosenthal and Strange, 2020). Our focus also contrasts with papers in which neighborhood peer effects and related social interactions sometimes contribute to crime (e.g., Billings et al., 2019).

[3] For related work on commercial activity and crime, see Hakim and Shachmurove (1996), Greenbaum and Tita (2004), Lee and Alshalan (2005), Bowes (2007), Stucky and Ottensmann (2009), Browning et al. (2010), Rosenthal and Ross (2010), Weterings (2014), Groff and Lockwood (2014), Hipp (2016).

[4] The 2018 NRF survey is at https://cdn.nrf.com/sites/default/files/2018-10/NRF-NRSS-Industry-Research-Survey-2018.pdf. Details of the 2018-2020 World Bank Enterprise survey are at https://www.enterprisesurveys.org. We based our 1.3% measure above on pooled loses from Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, Greece, Hungary, Ireland, Italy, Lithuania, Luxembourg, Netherlands, Poland, Portugal, Slovenia, and Sweden.

that NYC retailers lost roughly $1.38 billion to theft in 2018 while also spending $740 million on security.[5] Adding to these costs, in 2018 the New York City Police Department (NYPD) directed $1.60 billion of its budget for police patrols (New York City Council, 2018), a portion of which would have been intended to protect against property crime.[6] In comparison, estimates later in the paper suggest that under plausible conditions, a one standard deviation increase in block-level concentration of retail activity would reduce property crime by at least 8.5%. That represents a total savings among NYC retailers of roughly $117 million.

Our conceptual model relies on three key assumptions. The first is that spatial concentration of retail outlets enhances eyes on the street, amplifying the effect of public and private investment in anti-crime measures. This implies that spatial concentration should reduce crime. The second is that local government chooses an efficient level of public protection allowing for response from private business owners. We show that this implies that spatial concentration should reduce the equilibrium level of public and private investment in anti-crime measures.[7] Our third key modeling assumption is that prices for retail products, wholesale merchandise, and labor services are constant across neighborhoods. Crime deterrent effects of spatial concentration will then be capitalized into higher neighborhood retail rent.

These relationships motivate three regressions that we estimate, including the effect of retail spatial concentration on crime, the frequency of police stops (which we use as a proxy for investment in anti-crime measures), and commercial rent. In each case, multiple strategies are used to identify crime-deterrent effects. This begins with organizing point-specific data into very small neighborhood units and then controlling for extensive attributes that characterize a neighborhood. We also compare the incidence of crime and police stops in instances where the potential for eyes on the street differ, including daytime versus nighttime, and pre-pandemic versus COVID-19 lockdown.

---

[5] The U.S. Census reports that NYC sales in 2012 were $92.265 billion which is roughly $100,000 billion in 2018 dollars (https://www.census.gov/quickfacts/fact/table/newyorkcitynewyork/RTN130212).
[6] Details of the NYPD 2018 budget are at http://council.nyc.gov/budget/wp-content/uploads/sites/54/2017/03/056-NYPD-exec-1.pdf .
[7] This result also relies on the implicit assumption that criminals trade off potential return and costs in the spirit of Becker (1968). See Freedman and Owens (2016) for recent related empirical evidence.

Our identification strategy is fundamentally reduced form in nature, relying as described above on extensive point-specific data and a series of differencing exercises. In that sense, the nature of our question, data and research design do not allow for a pseudo natural experiment. Nevertheless, under plausible assumptions, we argue that our models likely understate crime deterrent effects. In part this is because our model suggests that in high crime neighborhoods establishments have an incentive to proactively concentrate in order to gain better protection. This would cause the crime-concentration pattern to be less negative. In this context, the main threat to our ability to identify a deterrent effect of retail spatial concentration on crime would be if there is an unobserved micro-geographic (at the city block level) attribute that is associated with less crime but higher concentration of retail activity. An example could be block-level income or property values, but these measures are included as controls for that reason.

An analogous argument applies to our ability to identify the effect of crime deterrence on retail rent. In this case, spatial concentration of retail activity is expected to increase retail sales because of shopping externalities whereas crime deterrence from spatial concentration reduces cost.[8] Our capitalization model is designed to take advantage of this difference. Starting from the firm's profit function, we derive a capitalization expression for which the dependent variable is non-inventory costs per dollar of sales. We then show that any advantages from shopping externalities reduce the dependent variable while cost savings from crime deterrence do the opposite. Controlling once again for other micro-geographic attributes of the immediate neighborhood, we then argue that shopping externalities likely cause our estimates to understate the effect of crime deterrence on retail rent.

We use geocoded point-specific data from New York City (NYC) to test the model predictions. In all cases, NYC is first divided into 0.2 by 0.2 mile grid cells, each of which is treated as a separate neighborhood. We then omit predominantly residential neighborhoods from our estimating samples in a

[8] For previous work on the effect of shopping externalities on sales, see Pashigian and Gould (1998), Gould et al. (2005), Koster et al. (2014), Johansen and Nilssen (2016), Clapp et al. (2019) and Koster et al. (2019). For evidence that crime and other local attributes affect commercial property values and/or rent, see Sivitanidou (1995), Lens and Meltzer (2016), and Rosenthal, Strange and Urrego (2021).

manner described later. For the crime and police stop models, activity is analyzed at the neighborhood level. For the capitalization models, rent is analyzed at the establishment level. The small size of our neighborhood units helps to reduce potential for unobserved factors and is consistent with evidence that the effect of crime in urban areas is highly localized.[9] Data on reported crimes and police stops are obtained from the New York Police Department (NYPD).[10] Establishment level data on input costs and sales are obtained from CompStak Inc. and Dun and Bradstreet (CompStak provides information on rent and space leased while Dun and Bradstreet provides information on employment and sales).

Results indicate that increasing neighborhood spatial concentration of retail activity from the 25th percentile to the 75th percentile reduces property crime by at least 9.4% and police stops by 12.1%.[11] Included in these models is an extensive set of neighborhood and building specific controls. Most important, this includes the level and composition of employment in the neighborhood, spatial concentration of non-retail industries, and cell phone data that further controls for business and non-business foot traffic. Additional controls include the presence of trees (as a proxy for amenities), presence of residential units, building age, building assessed value, and other neighborhood and building attributes.

Additional sample designs shed light on crowding and visibility as underlying mechanisms. We compare crime rates at night to those during the day and crime rates throughout 2018 to the first two weeks of the NYC COVID-19 lockdown (March 22nd – April 5th, 2020).[12] Crowds are diminished at night and during the lockdown for reasons unrelated to crime. Visibility is also diminished at night but would have been unaffected by the lockdown. In some models, we also measure spatial concentration in three

[9] In related work, Ellen et al (2013) report that crime in New York City increases on city blocks where a mortgage default has recently occurred, likely because of deleterious effects from undermaintained and/or vacant properties. Linden and Rockoff (2008) find that the presence of a registered sex offender in Mecklenburg County, North Carolina has a negative effect on residential property values within 0.1 miles. Pope (2008) obtains similar results for Hillsborough County, Florida. In all three studies, estimated effects attenuate rapidly with distance.
[10] The police stop data were collected as part of the NYPD policy of stop-question-frisk (SQF). The SQF policy was widely criticized up to roughly 2012 as contributing to discriminatory police behavior against minorities, prompting a sharp shift in policy implementation. We use police stop data from 2016-2018 to mitigate those concerns.
[11] In all cases, crime and police stop estimates are obtained from negative binomial count models.
[12] By late April 2020 up to 20% of the NYPD police force was out sick with COVID-19. This would have reduced the ability of police to patrol (https://www.cnn.com/2020/04/07/us/nypd-coronavirus-out-sick/index.html). Focusing on the first weeks of the lockdown avoids this issue.

different ways, based on spatial patterns of employment, the location of storefronts, and sales. We argue that the first measure is especially effective at capturing crowding effects, the second targets visibility, and the third is a placebo check having conditioned on the first two. Evidence from these strategies suggest that crowding and visibility both enhance crime deterrent effects from retail spatial concentration.

Results from our rent models confirm that crime deterrence is capitalized into higher local rent, reinforcing our core findings. For the average neighborhood, a one standard deviation increase in retail spatial concentration is associated with a 7.8% increase in expenditures on space and labor per dollar of sales. A corresponding estimate for wholesale establishments is smaller and serves as a robustness check. The absence of shoppers from warehouse facilities reduces the threat of shoplifting and allows for more aggressive protection measures that would discourage retail shoppers. Both effects should reduce the crime deterrent effect of retail spatial concentration on wholesale establishment rent.

To establish the results above, we proceed as follows. Our conceptual model is in Section 2. Section 3 describes the data and summary statistics. Section 4 presents results, and Section 5 concludes.

2. Model

This section has two parts, each of which highlight implications of the assumption that spatial concentration of retail activity amplifies the effect of eyes on the street. The first part shows that spatial concentration should reduce public and private investment in anti-crime measures. The second part derives an expression that allows us to estimate a lower bound on the degree to which crime-deterrent effects from spatial concentration are capitalized into higher neighborhood retail rent.

2.1 Investment in anti-crime measures

Suppose initially that all retail establishments are identical (this assumption is relaxed later). Each store in neighborhood j has inventory I_j, some of which is stolen while the rest is sold,

$$I_j = I^{stolen}(P_j) + I^{sold}(P_j) . \qquad (2.1)$$

Stolen inventory declines with protection against crime P_j, which is given by,

$$P_j = G_j Pu_j^{\alpha} Pr_j^{\beta} \quad \text{with } \alpha + \beta \leq 1 \;, \tag{2.2}$$

where Pu_j and Pr_j denote public and private expenditures on security, respectively, and G_j denotes spatial concentration of retail activity in neighborhood j. In (2.2), G_j increases protection as a Hicks neutral shift factor that amplifies the effect of Pu_j and Pr_j. This occurs because spatial concentration is expected to enhance the effect of eyes on the street as described earlier (e.g., Chang and Jacobson, 2017; McMillen et al., 2019; Gonzalez and Komisarow, 2020).[13]

Each establishment incurs the following expenses for private and public security measures,

$$exp_j = Pr_j + \frac{1}{N_j} Pu_j \;, \tag{2.3}$$

where N_j is the number of retail establishments in the community, each of which pays an equal tax share to support public protection. The price of public protection is then given by $1/N_j$ while the price of private protection is normalized to 1.

In the simplest setting, local government acts as a social planner and chooses Pu_j and Pr_j to minimize protection costs for each store while providing a socially efficient level of protection (P_j^*).[14] In Appendix A, we show that P_j^* and the equilibrium share of inventory lost to crime increase and decrease, respectively, with spatial concentration. This reflects the core modelling assumption that G_j deters criminal activity. The effect of G_j on public and private investment in anti-crime measures is obtained by taking first-order conditions of the local government's optimization problem and rearranging.[15] The efficient levels of Pu_j and Pr_j are given by,

[13] Spatial concentration could also facilitate collaboration with neighbors on protection measures as seems likely to occur in business improvement districts, BIDs (see, for example, Hoyt, 2005; Brooks, 2008 and Faggio, 2021). We consider this point in a robustness check and we conclude that BID presence does not affect our core results.

[14] Our results remain the same if local government seeks to maximize private sector profit by choosing the optimal level of Pu_j for a given Pr_j, and similarly, that the private sector chooses an optimal Pr_j for a given Pu_j.

[15] The corresponding Lagrangian is given by $\mathcal{L} = Pr_j + \frac{1}{N_j} Pu_j + \eta\left(P_j - G_j Pu_j^{\alpha} Pr_j^{\beta}\right)$.

$$Pu_j^* = \left[\frac{P_j^*}{G_j}\right]^{\frac{1}{\alpha+\beta}} N_j^{\frac{\beta}{\alpha+\beta}} \left(\frac{\alpha}{\beta}\right)^{\frac{\beta}{\alpha+\beta}} \qquad (2.4a)$$

$$Pr_j^* = \left[\frac{P_j^*}{G_j}\right]^{\frac{1}{\alpha+\beta}} N_j^{-\frac{\alpha}{\alpha+\beta}} \left(\frac{\alpha}{\beta}\right)^{-\frac{\alpha}{\alpha+\beta}} \quad . \qquad (2.4b)$$

From these expressions it is apparent that higher G_j reduces public and private investment in anti-crime measures. In the empirical work that follows, we do not observe Pr_j but we do observe police stops, which we use as a proxy for Pu_j.

2.2 Rent capitalization

2.2.1 Modeling assumptions

Additional modeling features are required to evaluate rent capitalization effects from crime deterrence associated with spatial concentration. We assume that retail product price, wage, and inventory cost (*p*, *w*, and *c,* respectively) are determined at the metropolitan level and do not vary across neighborhoods. Retailers sell q units of merchandise at price *p*, hire labor L at a wage *w*, rent space S at a rent per square foot r, and purchase inventory from wholesalers at a per unit cost c. Importantly, *r* varies across neighborhoods for two reasons. As above, the share of inventory lost to crime is assumed to shrink with neighborhood spatial concentration of retail activity, G_j. In addition, output q is produced using labor and space, the productivity of which are amplified by shopping externalities that draw more shoppers to a store, and which also increase with G_j. Both forces should cause neighborhood rent to increase but in a manner that causes our model to understate the capitalization effect of crime deterrence on neighborhood rent. We return to this point in the final portion of this section.

2.2.2 Homogenous establishments

Treating all establishments as alike, and collecting terms from above, profit for an establishment in neighborhood *j* is given by,

$$\pi = pq(L_j, S_j; G_j) - wL_j - r(G_j)S_j - c[1 + C(G_j)]q(L_j, S_j; G_j) \ , \qquad (2.5)$$

where $c\left[1+C\left(G_j\right)\right]$ is the cost of inventory for each unit sold. That cost increases with wholesale purchase price c and the share of inventory lost to crime, which we denote as $C\left(G_j\right)$. Setting $\pi = 0$ (with competitive markets), expenditure on space per dollar sold can be written as,

$$\frac{r(G_j)S_j}{pq(G_j)} = \theta - \frac{wL_j}{pq(G_j)} - \gamma(G_j) \quad , \tag{2.6}$$

where $\theta = \frac{p-c}{p}$ and $\gamma\left(G_j\right) = \frac{cC(G_j)}{p}$.

In (2.6), θ is the percentage markup of retail to wholesale price and is common across establishments in the metropolitan area. The term $\gamma\left(G_j\right)$ is a neighborhood-specific markup that allows for the cost of inventory lost to crime. The remaining term, $\frac{wL_j}{pq(G_j)}$, is labor cost per dollar sold. In our data we observe $\frac{L_j}{pq(G_j)}$ which is included as a control in some of the regressions. The coefficient on that term provides an estimate of w. In an alternate specification, we shift $\frac{wL_j}{pq(G_j)}$ to the left side of the equation and use earnings data for New York City from the U.S. Bureau of Labor Statistics (BLS) to measure labor cost per dollar of sales. Results from the two specifications are quite similar. Bearing this in mind, we rewrite (2.6) as,

$$\frac{r(G_j)S_j + wL_j}{pq(G_j)} = \theta - \gamma(G_j) \quad , \tag{2.7}$$

where the dependent variable measures non-inventory costs per dollar of sales.

2.2.3 Heterogeneous establishments

In this section we highlight three sources of heterogeneity that affect the dependent variable in (2.7). One is that companies belong to different industries, $k = 1, \ldots K$, each of which may have its own markup, θ_k. In the estimation to follow, we allow for this by including industry SIC 2-digit fixed effects. A second source of heterogeneity are neighborhood attributes apart from G_j that may also affect productivity. These terms are represented by z_j and include spatial concentration of non-retail economic activity, neighborhood level proxies for potential demand, and more. A third source is establishment-level skill that reduces S and L for a given q, denoted as θ_i, for i = 1, … I establishments. Collecting terms and suppressing the i subscripts on S and L to simplify, (2.7) becomes,

$$\frac{r(G_j)S_j+wL_j}{pq(G_j)} = \theta_k - \gamma\left(G_j\right) + bz_j + \theta_i \quad . \tag{2.8}$$

Our primary goal with (2.8) is to measure the effect of G_j on $\gamma\left(G_j\right)$. We do this in three ways. In the first approach, we estimate (2.8) using neighborhood fixed effects to measure $\gamma\left(G_j\right)$. The fixed effects are then regressed on G_j to summarize the average relationship between $\gamma\left(G_j\right)$ and G_j. A second, more general approach estimates (2.8) using Robinson's (1988) partial linear model. In this approach, $\gamma\left(G_j\right)$ is estimated nonparametrically while specifying a parametric structure for the other model terms. Results from both approaches indicate that $\gamma\left(G_j\right)$ is approximately linear in G. Partly for that reason, in a third approach we impose a linear approximation on $\gamma\left(G_j\right)$. Taking a first order Taylor expansion of $\gamma\left(G_j\right)$ around $\bar{G}$ and rearranging terms, expression (2.8) becomes, [16]

$$\frac{r(G_j)S_j+wL_j}{pq(G_j)} = \bar{\theta}_k - \gamma'(\bar{G})G_j + bz_j + \theta_i \quad , \tag{2.9}$$

where $\bar{\theta}_k = \theta_k - \gamma(\bar{G}) + \gamma'(\bar{G})\bar{G}$. Notice that if $\gamma\left(G_j\right)$ is linear then $\gamma(\bar{G}) = \gamma'(\bar{G})\bar{G}$ and $\bar{\theta}_k = \theta_k$ so that $\bar{\theta}_k$ equals industry markup. Also, and of primary interest, the coefficient on G_j in (2.9) measures the marginal effect of G on the cost of inventory lost to crime per dollar sold evaluated at $\bar{G}$.

In (2.9), it is worth noting that neighborhood attributes that enhance productivity and/or reduce crime deterrent costs are expected to increase non-inventory costs per dollar sold while a business owner's skill does the reverse. This is because local productivity advantages from $\gamma\left(G_j\right)$ and bz_j should be capitalized into higher local rent. Entrepreneur skill, as captured by θ_i, should instead increase profit. In the empirical work to follow, we proxy for θ_i using establishment age as it is well established that older companies tend to be more productive. As anticipated, non-inventory costs per dollar sold decline with establishment age but increase with neighborhood level retail spatial concentration.

[16] Expanding $\gamma\left(G_j\right)$ around $\bar{G}$, $\gamma\left(G_j\right) \approx \gamma\left(\overline{G}\right) + \gamma'\left(\overline{G}\right)\left[\overline{G} - G\right]$. Substituting into (2.8) yields expression (2.9).

2.2.4 Lower bound on rent capitalization from crime deterrence

As suggested earlier, the model above yields a lower bound on the degree to which crime deterrence arising from retail spatial concentration is capitalized into higher neighborhood rent. To see why, recall that shopping externalities are likely to increase with spatial concentration of retail activity (e.g., Gould et al., 2005; Koster et al., 2014; and Koster et al., 2019). For a given level of q, shopping externalities increase productivity by reducing the need for advertising and by improving the ability of store managers to anticipate flows of shoppers, enabling them to use space and labor more efficiently. These advantages will also be capitalized into higher local rent.[17]

Suppose now that crime deterrent effects are absent so that spatial concentration only affects profit through shopping externalities. Also, hold constant the level of space and labor used in production. Taking the derivative of the dependent variable in (2.9) with respect to G_j and manipulating, productivity advantages from shopping externalities cause input costs per dollar sold to shrink if the following condition holds:

$$\frac{r\prime(G_j)S_j}{r(G_j)S_j+wL_j} < \frac{q\prime(G_j)}{q(G_j)} \tag{2.10}$$

where $r'(G_j)$ and $q'(G_j)$ are derivatives with respect to G_j.

Note now that $wL_j > 0$ so that $\frac{r\prime(G_j)S_j}{r(G_j)S_j+wL_j} < \frac{r\prime(G_j)}{r(G_j)}$. Also, $\frac{r\prime(G_j)}{r(G_j)}$ and $\frac{q\prime(G_j)}{q(G_j)}$ are approximately equal to $\%\Delta r$ and $\%\Delta q$, respectively. A sufficient condition for (2.10) to hold is that productivity gains from shopping externalities have a smaller percentage effect on rent r than on sales q. Moreover generally, (2.10) will hold provided that $\%\Delta r$ is not substantially *larger* than $\%\Delta q$ since $\frac{r\prime(G_j)S_j}{r(G_j)S_j+wL_j} < \%\Delta r$. This condition is undoubtedly met. Shopping externalities, therefore, shrink the dependent variable in our

[17] Shopping externalities also have potential to increase q for a given retailer, requiring purchase of additional S, L and inventory. However, if production is approximately constant returns to scale, as seems likely, higher q would not affect the firm's profit margin or the dependent variable in (2.8). Instead, it is the efficiency gains and related productivity advantages from shopping externalities that enhance net profit.

capitalization expressions while crime deterrence has the opposite effect. For this reason, our model will understate the rent capitalization effect of crime deterrence.[18]

3. Data, Neighborhoods, and Summary Statistics

A complete list of the large number of variables and many data sources used to estimate the empirical models is provided in Appendix B. All measures focus on New York City for roughly 2018. For the crime and police stop models, the dependent variables and controls vary at the neighborhood level. For the capitalization models, the dependent variable is at the establishment level and some controls vary at that level while others are at the neighborhood level. All data are initially obtained as point-specific measures and then aggregated up to the neighborhood level as needed. Below we first describe how neighborhoods are measured. This is followed by a description of the data and summary measures.

3.1 Measuring neighborhoods and spatial concentration

3.1.1 Defining neighborhoods

For all of our models, we divide NYC into 0.2 by 0.2 mile grid squares. This corresponds to roughly two Manhattan blocks traveling east-west and three blocks traveling north-south (approximately a 4 to 7 minute walk). Grid squares are independent of administrative boundaries, and each is treated as a separate neighborhood. It is worth emphasizing that the grid squares are small enough to be relatively homogenous but large enough to allow for within-grid square variation in the spatial concentration of economic activity and other measures.

In total, 6,233 grid squares cover the five boroughs that make up NYC. Of these, 3,506 have an active commercial presence and are included in the estimating sample. The other 2,727 are predominantly residential and are omitted for that reason. Details are discussed later as are robustness checks which indicate that dropping residential neighborhoods does not affect our results.

[18] Note also that if capitalization of G into higher rent prompts retailers to substitute L for S, related cost savings will cause our model to further understate capitalization effects from crime deterrence.

3.1.2 *Measuring spatial concentration within grid squares*

We use the Getis-Ord statistic to calculate spatial concentration in a given grid square (Getis and Ord, 1992; Ord and Getis, 1995). This statistic is widely used for Hot-Spot analysis, especially for policing strategies that target hot-stop crime areas. To simplify exposition, a given target establishment is always indexed by *i* while all other establishments in our NYC sample are indexed by $e = 1, \ldots n$. Using our prior notation, the Getis and Ord expression for G_i is given by,

$$G_i = \frac{\sum_{e=1}^{n} \omega_{ie} x_e - \bar{X} \sum_{e=1}^{n} \omega_{ie}}{\sqrt{\frac{\sum_{e=1}^{n} x_e^2}{n} - \bar{X}^2} \sqrt{\frac{\left[n \sum_{e=1}^{n} \omega_{ie}^2 - \left(\sum_{e=1}^{n} \omega_{ie}\right)^2\right]}{n-1}}} \quad . \tag{3.1}$$

In this expression, x_e is employment at establishment e and $\bar{X}$ is the average size of an establishment throughout our NYC sample.

A key feature when implementing (3.1) is to specify a function for ω_{ie}, the weight placed on employment as distance, d_{ie}, increases from establishment *i*. We adopt the following weight function:

$$\omega_{ie}(d_{ie}) = \begin{cases} 1, & \text{if} \quad d_{ie} \leq 250 \\ 1/(d_{ie} - 250)^{0.7}, & \text{if} \quad 250 < d_{ie} \leq 1{,}000 \\ 0, & \text{if} \quad d_{ie} > 1{,}000 \end{cases} \tag{3.2}$$

This function sets ω_{ie} to 1 for all establishments within 250 feet of *i*. For distances between 250 to 1,000 feet from *i*, ω_{ie} is assumed to decline with distance at rate $1/(d_{ie} - 250)^{0.7}$, where the exponent 0.7 was chosen to set ω_{ie} to roughly 1% at 1,000 feet. Beyond 1,000 feet, ω_{ie} is set to zero. Measured in this fashion, the weight function will often apply positive weight to employment beyond the border of a grid square. Results were also robust to alternate reasonable specifications of *G*.[19] Note further that specified as above, G_i is measured separately for each establishment and varies within a given neighborhood.

[19] Results were similar for different thresholds from 250 to 1,000 feet in expression (3.2). Estimates were also robust to exponents of 0.5, 1 and 2 that govern the rate of decay in the inverse distance portion of ω_{ie}.

For the crime and police stop models, the unit of analysis is the neighborhood. For that reason, we aggregate G_i to the neighborhood level in those models and also normalize G across neighborhoods to simplify interpretation. The resulting measure $\tilde{G}_j$, is formed as,

$$\tilde{G}_j = \frac{1}{sd(\bar{G}_j)}\left(\bar{G}_j - \frac{1}{m}\sum_{j=1}^{m}\bar{G}_j\right) \quad . \tag{3.3a}$$

In this expression, $\bar{G}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} G_i$ is the average level of spatial concentration in neighborhood *j* (with n_j establishments) while *m* is the number of grid squares in the estimating sample. Measured as above, $\tilde{G}_j$ is positive if the average level of spatial concentration in neighborhood *j* is high relative to the typical neighborhood. Also, a 1 unit increase in $\tilde{G}_j$ represents a 1 standard deviation increase in spatial concentration across grid squares.

For the rent capitalization models, the unit of analysis is the individual establishment. In those models, we normalize G_i in a fashion analogous to above while allowing the concentration measure to vary across establishments within individual neighborhood grid squares. Specifically, we form,

$$\tilde{G}_i = \frac{1}{sd(G_i)}\left(\bar{G}_i - \frac{1}{I}\sum_{i=1}^{I}\bar{G}_i\right) \quad . \tag{3.3b}$$

where *I* is the number of establishments throughout the entire estimating sample.

In the estimation to follow, our primary measure of $\tilde{G}$ is based on the spatial distribution of employment and is denoted as $\tilde{G}_{Emp}$ (we drop the *j* and *i* subscripts for convenience). This captures a combination of effects from crowding and visibility (the ability to observe multiple storefronts at once). As mentioned in the Introduction, in some models we also add a measure of $\tilde{G}$ based on the spatial concentration of storefronts, $\tilde{G}_{Stores}$, as this directly targets storefront visibility. In this instance, $\tilde{G}_{Stores}$ is measured as if there is only one worker at each store and is formed as:

$$\tilde{G}_{\text{Ret,Stores}} = \sum_e \omega_{ie}(d_{ie}) / n(d_{ie}), \tag{3.4}$$

where n_i is the number of establishments within a given distance of store *i* (d_{ie}), ω_{ie} is defined as in (3.2), and $\tilde{G}_{Stores}$ is normalized as in (3.3a). As a placebo check, we also consider a third measure of

$\tilde{G}$ based on sales, denoted as $\tilde{G}_{Sales}$ (calculated in the manner as $\tilde{G}_{Emp}$ except using sales). Controlling for $\tilde{G}_{Emp}$ and $\tilde{G}_{Stores}$, we do not expect $\tilde{G}_{Sales}$ to affect eyes on the street.

When measuring the level and composition of neighborhood employment, we use all establishments in the neighborhood. When measuring $\tilde{G}_{Emp}$ however, we use only single site firms which account for 95 percent of all establishments in NYC. We do this to better target external economies of scale that arise from clustering individual companies together. The presence of a single big-box retail chain outlet, for example, could cause $\tilde{G}_{Emp}$ to be high without any actual spatial concentration of establishments. As a robustness check, we also included establishments at multi-site firms when measuring $\tilde{G}_{Emp}$. Results (reported later) are similar.[20]

3.2 Data

A complete list of measures used in our regression models and their sources is provided in Appendix B. Below we comment on different groupings of the many types of data used in the analysis.

3.2.1 Property crime

The New York Police Department (NYPD) provides data on all criminal complaints since 2006. In most instances, we use data for 2018. For each crime, this includes the date, time, precise location, and type of crime. Property crime includes petit larceny, grand larceny, burglary, theft of services and fraud. We aggregate these crimes together for our core models but estimate separate models in other instances,

[20] It is worth noting that single-site companies are small compared to establishments that belong to multi-site firms. Among all industries combined, the number of workers in NYC at the 75th, 95th and 99th size percentiles are 3, 10, and 40 for single-site companies, and 26, 150 and 500 for establishments of multi-site firms. Reflecting these differences, single-site companies also often cannot afford the cost of private loss prevention measures. In an NBC news report (December 2, 2021), for example, Andrew Dimian, CFO of Omni private security services comments: "A lot of small businesses have been contacting us, but they just can't afford having a [security] guard there … About 90 percent of the businesses that ask can't afford it." See: https://www.nbcnews.com/business/business-news/small-businesses-no-easy-way-fight-smash-grab-robberies-rcna7337.

in addition to models for robbery and auto theft which also entail theft of property. Nighttime crime, as highlighted in some models, is defined as crime after 10 pm and before 5 am.

We drop crimes committed on a bus, subway or at a subway station as the location of the event may not be accurately coded. We also drop all complaints where the perpetrator left the scene before fully committing the offense (less than 2% of all events), and crimes that extend beyond one day, as with kidnapping and/or hostage situations. This leaves 101,896 property crimes in the analysis to follow.

3.2.2 Police stops

Police stop data were obtained from the New York Police Department (NYPD) and are reported as part of the Stop-Question-Frisk (SQF) policy. We pool stops from 2016-2018 as this helps to ensure a large enough number of stops to obtain reliable estimates.[21] These years also post-date a well-documented shift in how the SQF policy was implemented, a primary goal of which was to better target criminal behavior. SQF stops peaked in 2012 at over 500,000 (Evans et al., 2014) and were associated with widespread allegations of police bias against African Americans, evidenced in part by low arrest rates among African Americans stopped under the policy. By 2016, SQF stops had shrunk to just over 10,000 per year (12,053, 11,204 and 11,008 in 2016, 2017 and 2018, respectively) while arrest rates increased sharply (e.g., Urrego, 2023), indicative of better targeting of criminal activity.[22]

SQF data pertain only to pedestrian stops and include the date, time, and location of each stop. Also reported is whether an arrest was made and if so for what type of crime. We drop any stops prompted by a 911 call or which were associated with an ongoing investigation. Instead, we retain only

[21] Restricting police stop data to 2018 did not affect the results but increased the share of grid squares with zero stops from 50% to 75%. Using Zero-Inflated Negative Binomial models also did not affect the results.
[22] Studies based on early years of SQF found that African Americans and Hispanics were stopped at a higher rate than their white counterparts, even after controlling for neighborhood racial composition and criminal activity by race (Gelman et al., 2007; Ridgeway, 2007; Hanink, 2013; Evans et al., 2014; Ferrandino, 2018;). However, MacDonald and Braga (2019) show that racial patterns associated with SQF appear to have decreased in more recent years. Also worth noting, other studies have found that SQF stops reduce crime (Weisburd et al., 2016; Wooditch and Weisburd, 2016; MacDonald et al., 2016; Rosenfeld and Fornango, 2017; Ferrandino, 2018).

stops in which police were acting proactively. This reduces the number of stops in the analysis, from 34,265 (for the 2016-2018 period) to 7,277.

3.2.1 Employment, sales, and industry

Dun & Bradstreet provides information on more than one million establishments in the New York City area. For each establishment, we observe employment, sales, industry code, address, and latitude-longitude coordinates. The data were downloaded from the Syracuse University library (which has a site license) between October 2018 and February 2019. These data were used to construct a variety of key neighborhood-level controls in addition to the dependent variable for the capitalization model.

3.2.2 Commercial rent

Commercial lease data were obtained from CompStak Inc. and are used to estimate the rent capitalization models. For each lease, CompStak reports effective rent per square foot of space leased, location of the lease (including street address and latitude and longitude), and tenant name. Our lease sample includes over 60,000 leases in NYC that were executed up to December 2019. We match these data at the establishment level with the D&B data using information on tenant name, street address, and latitude/longitude coordinates. In total, we are able to reliably match almost 50% of the CompStak leases. Of the leases that were matched, 4,000 are classified as retail establishments in D&B with a primary SIC code 52-59. Of these, more than half include missing information on sales, employment or space leased, measures needed to estimate the models in expressions (2.6), (2.8) and (2.9). This leaves us with roughly 1,600 observations for the retail rent capitalization portion of the analysis. An additional roughly 550 matched observations are used to estimate analogous models for wholesale establishments.

3.2.3 Additional neighborhood controls

An extensive set of local attributes based on point-specific data were derived from various New York City government agencies and coded up to the neighborhood level. This includes data from the

MapPLUTO 18v2 map which is produced by the New York City Department of City Planning and the Department of Finance. The map provides detailed information on the attributes of each tax lot in NYC, including building attributes, zoning, tax assessments, and other lot specific characteristics. Additional neighborhood level data were obtained from the NYC Department of Health and Mental Hygiene, Fire Department, Department of IT and Telecommunications, The NYC Community Air Survey, the NYC Open Data portal, and the 2015 Tree Census conducted by the NYC Department of Parks and Recreation. Drawing on these sources, our primary models control for neighborhood-level measures of the share of residential units in the grid square, total number of trees in the grid square, whether the grid square overlaps multiple police precincts, average age of buildings, average assessed value of buildings, and average sales per worker for single-site establishments (including companies in all industries).

Also included in our primary models is a measure of local foot traffic based on cellphone data from SafeGraph. SafeGraph defines over 110,000 Points of Interest (POI) in New York City and measures visits to each POI using cellphone GPS information combined with information on building footprints and other relevant information (e.g., store open hours). In cleaning these data, we first calculate the average number of monthly visitors to each individual POI during 2018. For each grid square, we then average monthly visits across POI within a grid square. Along with neighborhood level employment, these measures provide considerable information on the level of activity in a neighborhood.[23]

3.3 Neighborhood sample coverage and summary statistics

The set of neighborhoods included in our estimating sample is implicitly determined by our model specification. For most of the crime and police-stop models we include separate measures of $\tilde{G}_{Emp}$ for retail, finance, service and manufacturing, where each industry is defined based on its SIC code (52-59 for retail, 60-67 for finance, 70-89 for the service sector, and 20-39 for manufacturing). This allows us

[23] We also calculated spatial G measures for POI. Correlation between that measure and spatial concentration of retail employment was 8%, indicating that the measures capture different information. Including spatial concentration of POI visits had no effect on the other model coefficients and was dropped to simplify discussion.

to compare the effect of spatial concentration of retail activity to that of other industries. An important finding is that spatial concentration of retail has a more notable effect. This approach also limits the estimating sample to neighborhood grid squares in which all four highlighted industries have at least some presence. Table 1a compares the attributes of the included and omitted neighborhood grid squares.

As is apparent in the table, omitted grid squares are predominantly residential areas. On average, grid squares included in the estimating sample have 239 establishments with total employment equal to 1,542. Among grid squares omitted from our estimating sample, the corresponding values are 40 and 214, respectively. As a robustness check, in the estimation to follow we report estimates in which only retail and non-retail activity (as a single category) is highlighted. This increases the number of neighborhoods in the model from 3,506 to 5,461. To anticipate, results are robust.

Figures 1a and 1b display heat maps of the spatial patterns of total retail employment and $\tilde{G}_{Retail,Emp}$ for the 3,506 grid squares included in our primary regressions. Grid squares not included are white. Notice that included grid squares are drawn from throughout the NYC area and include nearly all of Manhattan, the employment center for NYC. Also apparent, while retail employment is heavily concentrated in a band extending south from Central Park, $\tilde{G}_{Retail,Emp}$ varies more widely. Figure 1c zooms in on the area south of Central Park and overlays individual establishment location on top of retail spatial concentration, with larger circles for establishments with more employment. This figure shows that there is considerable spatial variation in $\tilde{G}_{Retail,Emp}$ even after controlling for the size of nearby retail establishments. Indeed, over the entire sample of 3,506 neighborhood grid squares, the correlation between $\tilde{G}_{Retail,Emp}$ and neighborhood retail employment is just 28%. Retail employment and $\tilde{G}_{Retail,Emp}$ contain different information as is also apparent in the regression models that follow.

Table 1b provides additional summary measures for the 3,506 grid squares included in the primary estimation. In 2018, a grid square experienced an average of 29 property crimes, 63% of which were petit larcenies, with the total number of property crimes equal to 101,896. The number of police stops used in the analysis is smaller, just 7,277. Note also that roughly half of grid squares experience no police stops whereas the number of grid squares that reported zero property crime is below 2%. Because the crime and police stop data are count measures, and to allow for zeros, we estimate both the police stop and crime models using a negative binomial specification. This model is well suited to sample distributions such as ours for which the variance of the outcome measures exceeds their means.[24]

In Table 1b, notice that the service industry accounts for the highest share of employment among the industries highlighted (46%), followed by retail (19%), finance (7%), and manufacturing (5%). Observe also that of the four industries highlighted, retail employment is the most spatially concentrated based on both the median and 75th percentile values across the sample of neighborhoods.

Also in Table 1b, the ratio of non-inventory costs per dollar sold (including labor costs) as well as the ratio of the leased space per dollar of sales often exceed one. This is to be expected given the heavy reliance of small businesses on financing. The pattern is also consistent with a high failure rate of retail establishments, presumably because many establishments are unable to generate sufficient revenue to cover costs.[25]

4. Results

4.1 Property crime and police stops

4.1.1 Core estimates

Table 2 reports estimates of the effect of employment-based spatial concentration of retail activity on property crime and police stops, $\tilde{G}_{Emp}$. As noted earlier, these estimates are obtained from negative binomial count models that address zeros in the data. Marginal effects evaluated at the mean of the full set of control measures are reported.[26] Recall that $\tilde{G}_{Emp}$ is normalized to have mean zero and standard

---

[24] We also estimated OLS regressions for both police stops and property crime, setting the dependent variables to log(X+1) with X suitably defined in each instance. Results were similar.

[25] Based on data from the US Bureau of Labor Statistics, roughly 20% of small businesses fail in their first year, 50% by their fifth year, and 70% in their first ten years.

[26] Marginal effects from a negative binomial regression can be calculated using the expression, $\exp(\beta_x \Delta x)$, where Δx represents the change in a control variable and β_x is its corresponding coefficient. This expression gives the change in counts. β_x also approximates a semi-elasticity because a one unit change in x represents a change in the log count of the dependent variable equal to β_x.

deviation of 1 so that a one-unit change in $\tilde{G}_{Emp}$ equals one standard deviation across neighborhoods. Columns 1-5 present estimates of property crime while columns 6-10 repeat the estimation with police stops as the dependent variable.

For both the crime and police stop models, the first two columns demonstrate that our primary estimates are robust to the omission of predominantly residential neighborhoods from our estimating sample. In these columns, we control for just three measures, aggregate neighborhood employment, retail share of grid square employment, and spatial concentration of retail employment. In column 1 (for crime) and column 6 (for police stops), the sample includes all 5,461 neighborhoods with at least some presence of both retail and non-retail employment. In columns 2 and 7, the sample is restricted to the 3,506 neighborhoods in which retail, finance, manufacturing, and service sector establishments are all present. For both the crime and police stop models, the coefficients on total employment and retail share of employment are very similar for the two different samples. The coefficients for retail spatial concentration become smaller when shifting to the more restricted sample but remain significant. For crime, the respective coefficients for the two samples are -0.097 and -0.061, while for police stops the corresponding values are -0.150 and -0.087.[27]

Columns 3 (for crime) and 8 (for police stops) add in a large number of neighborhood-level controls. Most important, these include employment shares and spatial concentration for the non-retail industries. Also included is the log of grid square sales per worker (based on single site establishments), the log number of trees in the neighborhood, average age of the buildings, log of the average building assessed value, whether the grid square overlaps with more than one police precinct, and the share of residential building units from among all buildings in the grid square. These columns also include average monthly visits to POI in the grid square to further control for the level of neighborhood activity. Comparing back to the estimates in columns 2 (crime) and 7 (police stops), respectively, it is noteworthy

[27] Analogous results are obtained when controls are added for the service sector which is present in all neighborhoods where retail is present.

that the addition of so many neighborhood-level controls has little effect on the coefficients on retail employment and spatial concentration. Once again, the estimates of interest are robust.

The magnitude of the coefficients in columns 3 and 8 is also of interest. Point estimates suggest that doubling grid square aggregate employment is associated with an increase in property crime and police stops of 55% and 35%, respectively. The corresponding effects from a doubling of visits to POI are 52% for property crime and 89% for police stops. These estimates confirm priors that property crime and police stops tend to be higher in neighborhoods with more activity.

Also noteworthy, the coefficients on retail share of employment in both the crime and police stop models are large, positive, highly significant, and much larger in magnitude than for the other industries (finance, manufacturing and service): a 1 percentage point increase in retail share of employment is associated with a 2.3% increase in property crime and a 2% increase in the number of police stops. These patterns confirm that retail activity, with its lucrative inventory, has an especially large effect on property crime and police activity.

Observe next that in column 3, a one unit increase in $\tilde{G}_{Ret,Emp}$ – equal to a one standard deviation increase in spatial concentration of retail employment across neighborhoods – is associated with an 8.5% decline in crime. In contrast, coefficients on spatial concentration for the other industries (finance, manufacturing, and service) are much smaller and mostly not significant. This pattern suggests that spatial concentration of retail activity has a particularly important effect on eyes on the street, reducing the cost of protection and causing equilibrium levels of crime to decrease.

The same pattern is present in column 8 for the police stop model with the exception that spatial concentration of service employment also has a similarly negative effect as for the retail sector. Increasing $\tilde{G}_{Ret,Emp}$ by 1 standard deviation decreases police stops by 11%. This supports the view that because spatial concentration reduces crime it also reduces investment in protection measures.[28]

[28] We recognize that police may limit pedestrian stops when their own behavior is more readily observed (see Owens, 2019, 2020, for related discussion). That concern is mitigated, however, by evidence in column 8 that total neighborhood employment and foot traffic to POI both have strong positive effects on police stops. Also, non-retail

A final comment concerns the potential for our estimates to understate crime-deterrent effects of spatial concentration. Our model implies that businesses have a financial incentive to spatially concentrate in high crime neighborhoods in order to enhance protection. To the extent that companies behave in this manner, that will cause the spatial concentration coefficients in Table 2 to be less negative, understating the crime deterrent effect of spatial concentration.

4.1.2 Additional robustness checks

Two additional sets of robustness checks are presented in Table 2. In columns 4 and 9 (for crime and police stops, respectively) we repeat the models specified in columns 3 and 8 but adjust how $\tilde{G}_{Emp}$ is measured. In these columns, we include establishments belonging to both single site and multi-site firms when measuring spatial concentration. Results are largely the same as for the specifications in columns 3 and 8. The primary difference is that the coefficient on $\tilde{G}_{Ret,Emp}$ is somewhat smaller for crime (in column 4) and larger for police stops (in column 9), but the qualitative pattern is the same.

In columns 5 and 10 (also for crime and police stops, respectively) we use a different neighborhood design that requires a brief description. In these models, each neighborhood is formed as a 3-by-3 configuration of the grid squares previously defined with the center square serving as a reference point to identify the neighborhood. Constructed in this fashion, neighborhoods overlap because a separate neighborhood is specified around each grid square. $\tilde{G}_{Emp}$ is then measured for a given 9-block neighborhood by forming:

$$\tilde{G}_{Emp,i} = \sum_{j=1}^{9} \left(\frac{E_{ij}}{\sum_{j=1}^{9} E_{ij}} \right)^2 \quad . \tag{4.1}$$

where $\tilde{G}_{Emp}$ equals 1 if all employment is in a single grid square and 1/9 if employment is spread equally across all 9 squares.[29] As a further robustness check, we used the same design as above but restricted our

employment always has a near zero effect on stops as does spatial concentration for FIRE and manufacturing. These patterns would not be anticipated if police shy away from making stops when others are present.

[29] Sample size for the 9-grid square neighborhoods is larger than for the individual 0.2 square mile neighborhoods used in other columns of Table 2. This is because the larger area used to define a neighborhood (nine grid squares)

estimating sample to non-overlapping neighborhoods. Results were quite similar to those in columns 5 and 10. Comparing estimates in columns 5 and 10 to those in columns 3 and 8, it is important to recognize that the scale of $\tilde{G}_{Emp}$ differs with the alternate neighborhood design. This is because we do not normalize $\tilde{G}_{Emp}$ to have unit standard deviation in this instance, which accounts for the much larger coefficients on $\tilde{G}_{Emp}$. Bearing that in mind, the qualitative pattern remains largely unchanged. The primary difference is that spatial concentration of service sector activity has a stronger deterrent effect relative to columns 3 and 8, but still not as large as estimates for the retail sector.

Summarizing, for each of the alternate model designs, the central results are robust. This includes that (i) retail employment share has a disproportionately positive effect on crime; (ii) that retail spatial concentration deters crime and reduces police stops; and (iii) that spatial concentration among non-retail industries has much less effect on crime and police stops.[30]

4.1.3 Mechanisms: crowding and visibility

Tables 3 and 4 report estimates from alternate specifications that are designed to shed light on crowding and visibility as mechanisms that may contribute to crime deterrent effects of spatial concentration. Table 3 compares estimates for different sample periods and types of crime. In Table 4 we compare estimates for the three different measures of spatial concentration described earlier, $\tilde{G}_{Ret,Emp}$, $\tilde{G}_{Ret,Stores}$ and $\tilde{G}_{Ret,Sales}$. In both tables, we report only estimates for crime using the same specification

reduces the number of instances in which all four highlighted industries are not present (retail, services, finance, and manufacturing) as described for the initial columns of Table 2.

[30] Two additional sets of models not reported in Table 2 were estimated as further robustness checks. The first added 25 additional neighborhood level controls to the models specified in columns 3 and 8. These included controls for establishment attributes (age and risk profile of neighborhood establishments), zoning (historic district, FAR restrictions), distance to important sites (e.g., subway stations, public parks), other neighborhood features (e.g., presence of rats, problems with light shine), and tax exemptions on buildings. Results were largely robust but are not preferred given concerns about collinearity among the many regressors and the possibility that some controls could be endogenous. In a different robustness check, we modified columns 3 and 8 to include controls for the portion of a grid square that belongs to a Business Improvement District (BID). This is because BIDs may pool resources and invest in local security measures (e.g., Faggio, 2021). Three patterns emerged: (i) estimates of the relationship between BID presence and crime were sensitive to how BID presence was specified; (ii) BID presence had no discernible relation to police stops; and (iii) controlling for BID presence had little effect on the coefficients for spatial concentration.

as in column 3 of Table 2, in part because there are too few observations on police stops in some instances to obtain reliable estimates. To conserve space, only the coefficients on retail share of employment and retail spatial concentration are reported.

Focusing first on Table 3, note that columns (1) and (2) are based on crime at all hours of the day and sample periods. Columns (3) and (4) pertain to crime during daytime hours, while columns (5) and (6) refer to crime at night. Columns (7) and (8) focus on crime during the first two weeks of the COVID-19 lockdown in New York City (for all hours of the day), March 22 to April 5 in 2020. Observe also that the upper rows in Table 3 correspond to different types of property crime while the bottom rows correspond to robbery and auto theft.[31]

When considering the patterns in Table 3, crowding is reduced at night and during the lockdown.[32] Visibility is reduced but not eliminated at night because of street lighting and would have been fully viable during the lockdown. Petit larceny is often associated with shoplifting that can only occur when a store is open. Grand larceny carries more serious penalties and is often associated with night-time break-ins that include burglary as part of the offense. It should also be noted that half of robberies in the NYC data occur on the street and not in a building; because of the small number of robberies, we do not attempt to decompose robberies by place of occurrence.

Focus now on the differences between daytime and nighttime patterns for crime. For all types of property crime aggregated together (the top row), at night the coefficient on retail share of employment is reduced by roughly 25% but remains large and highly significant. This suggests that inventory continues to attract criminal activity though thieves will need to break into stores that are closed, adding burglary to their crime. A different pattern is present for retail spatial concentration. The coefficient on that measure shrinks by roughly 95% at night and is no longer significant. This suggests that deterrent effects from

[31] Of the crimes highlighted in Table 3, grand larceny and petit larceny differ based on the value of merchandise stolen and can occur with or without breaking into a store. Break-ins are a defining feature of burglary. Theft of services often occurs when patrons leave a restaurant or hotel without paying for services. Fraud includes using a stolen credit card, forging signatures on a check, etc. Robberies occur when a victim is physically threatened.
[32] SafeGraph cell phone data indicate that foot traffic in NYC fell by 60% in April 2020.

crowding that arise from spatial concentration of retail activity are greatly reduced at night. These patterns are present for all categories of property crime.

For robbery, estimates are similar to property crime during the day but weaken only slightly at night: both retail employment and spatial concentration continue to have strong effects of the anticipated signs (positive and negative, respectively). Because crowding is mostly absent at night, visibility may remain a viable mechanism in helping to prevent robbery at night, possibly because roughly half of robberies in the NYC data occur on the street and not in a building.

Auto Theft exhibits a similar pattern as property crime during the day but the coefficients on retail employment and spatial concentration are noticeably smaller. At night, however, the coefficient on spatial concentration increases in magnitude, is negative, and strongly significant. If the nighttime concentration of parked cars is higher closer to retail establishments (which includes bars and restaurants), then a similar explanation as for robberies may apply. Police patrols and pedestrians may observe more cars at once where vehicle density is higher, and this may help to deter auto theft.

Columns (7) and (8) of Table 3 provide analogous estimates for the first two weeks of the spring 2020 COVID-19 lockdown in New York City. The dominant patterns are the same as for nighttime crime in column 6. For property crime, spatial concentration has much less effect relative to 2018 (in column 2) but spatial concentration has a similar deterrent effect on robberies as for the pre-pandemic period.

Bearing in mind that crowding is sharply reduced at night and during the lockdown, the patterns in Table 3 suggest that crowding and visibility have different effects on the different types of crimes considered in these tables. For Petit Larceny and Theft of Services/Fraud, the effect of retail spatial concentration is small and not significant both at night and during the lockdown. This along with other patterns suggests that crowding is an important deterrent of these crimes beyond simply having a store or restaurant open for business. For robbery, grand larceny (during the pandemic lockdown), and auto theft, the patterns suggest that visibility also acts as a deterrent.

Consider next Table 4 which includes controls for $\tilde{G}_{Emp}$, $\tilde{G}_{Stores}$, and $\tilde{G}_{Sales}$ for each of the four highlighted industries (retail, finance, manufacturing and service). Conditioning on all three measures at once, $\tilde{G}_{Emp}$ primarily targets crowding, $\tilde{G}_{Stores}$ proxies the potential to observe multiple storefronts from a single location (a feature of visibility), and $\tilde{G}_{Sales}$ is more of a placebo check as we have little reason to expect it to affect crime having conditioned on the other model controls.

We begin with Panel A of Table 4 which displays correlation coefficients for the three measures of spatial concentration for the retail sector. As would be anticipated, correlation between $\tilde{G}_{Ret,Emp}$ and $\tilde{G}_{Ret,Sales}$ is high, 54%. Correlation between $\tilde{G}_{Ret,Emp}$ and $\tilde{G}_{Ret,Stores}$, however, is just 14%, and correlation between $\tilde{G}_{Ret,Sales}$ and $\tilde{G}_{Ret,Stores}$ is 15%. These summary measures confirm that $\tilde{G}_{Ret,Emp}$, $\tilde{G}_{Ret,Stores}$ and $\tilde{G}_{Ret,Sales}$ contain different information.

Consider now Panel B of Table 4 which presents estimates of the property crime model for the same mix of time periods as in Table 3. Observe also that each row now corresponds to a different time-period regression with coefficients arrayed across columns. These include retail share of employment and the three different measures of retail spatial concentration, $\tilde{G}_{Ret,Emp}$, $\tilde{G}_{Ret,Stores}$ and $\tilde{G}_{Ret,Sales}$. Coefficients on the other model controls are suppressed to conserve space.

The patterns in Table 4 reinforce those in Table 3 for property crime. Crowding, as proxied by $\tilde{G}_{Ret,Emp}$, clearly deters property crime during the day and in the lockdown. But as in Table 3, the effect of $\tilde{G}_{Ret,Emp}$ is reduced at night when crowds are mostly absent. This pattern once again supports the idea that crowding enhances the crime deterrent effect of eyes on the street.

Notice also that visibility, as proxied by $\tilde{G}_{Ret,Stores}$, deters property crime during the day, at night, and during the COVID-19 lockdown. The corresponding coefficients are highly stable and significant, ranging between -0.12 and -0.14 across sample periods. This pattern is suggestive that visibility improves the ability of individuals (e.g., police or other individuals) and/or security cameras to observe potential criminal activity, enhancing crime deterrence.

A last point to note in Table 4 is that the coefficients on $\tilde{G}_{Ret,Sales}$ are always small and not significant. This was anticipated.

4.2 Rent capitalization

Our final models consider retail rent capitalization using matched establishment-level data. As described earlier, our estimates likely yield lower bound measures of the capitalization effect of crime deterrent effects from neighborhood spatial concentration.

We begin by estimating (2.7) and (2.8) from Section 2. Figure 2 displays scatter plots of the estimated fixed effects against spatial concentration of retail employment on the horizontal axis. Panel A omits controls for other factors (as in (2.7)) while Panel B controls for additional neighborhood and establishment attributes that may affect the dependent variable (as in (2.8)). In both panels, the scatter plots clearly increase with spatial concentration of retail employment. This confirms that benefits from crime deterrence are capitalized into higher rent.

Figure 3 displays an alternate set of estimates of $\gamma(\tilde{G}_{Ret,Emp})$ using Robinson's (1988) two-step partial linear model drawing on the semipar routine in Stata (Verardi and Debarsy, 2012). This estimates the $\gamma(\tilde{G}_{Ret,Emp})$ function non-parametrically with optimal smoothing. As before, Panel A does not allow for other factors while Panel B controls for additional neighborhood and establishment-level attributes. In both panels the gamma function is clearly increasing with retail spatial concentration and the confidence bands are narrow relative to the overall pattern. It is also evident that $\gamma(\tilde{G}_{Ret,Emp})$ is approximately linear in $\tilde{G}_{Ret,Emp}$. This last observation supports our remaining empirical exercise.

Table 5 reports estimates of expression (2.9) in which we impose a linear approximation on the relationship between $\gamma(\tilde{G}_{Ret,Emp})$ and $\tilde{G}_{Ret,Emp}$. Panel A uses rent per square foot per dollar sold as the dependent variable and includes labor per dollar sold as a control measure (as in expression (2.6)). Panel B shifts labor costs based on census data to the left side of the equation and uses non-inventory costs per dollar sold as the dependent variable as in Figures 1 and 2. Additional estimates in Panels C and D are based on 538 wholesale establishments and serve as a robustness check. In these panels, rent, labor costs and sales are all specific to the wholesalers in the sample, but spatial concentration is still measured using the same industries as before. As noted earlier, wholesalers are less prone to shoplifting and may also be more able to adopt aggressive anti-crime measures that would discourage retail shoppers. For these reasons, wholesale rent should be less sensitive to crime deterrent effects from retail spatial concentration.

In each panel, five models are presented with increasing numbers of controls for neighborhood and establishment-level attributes. We control for spatial concentration of other industries in column 2. In column 3 we add in the neighborhood level control for monthly visits to POI. Column 4 accounts for police precinct fixed effects, and column 5 adds in fixed effects for establishment age. A quick review across columns confirms that retail spatial concentration is associated with higher retail and wholesale rent. The magnitude of the spatial concentration effects also tends to shrink with further controls.

Focus now on column 5 for Panels A and C. The coefficient on employment per dollar of sales in Panel A for retail establishments is roughly \$38,000. The analogous estimate in Panel C for wholesale establishments is \$70,000. In comparison, for 2019, the office of the New York State comptroller reports that average retail earnings per worker in Manhattan were \$59,400 and for all of NYC \$46,600. For wholesale industry workers the US Bureau of Labor Statistics reports average 2019 earnings in NYC of \$88,000.[33] Our estimates are close to these values.

Consider now Panel B where we measure labor costs per dollar sold directly by combining the BLS measure of average earnings among NYC retail workers with Dun and Bradstreet data on employment and sales. That term is moved into the dependent variable as noted above and as in expression (2.7). The estimated effect of spatial concentration of retail employment in Panel B is very similar to the corresponding estimate in Panel A. A one standard deviation increase in spatial concentration of retail increases non-inventory costs per dollar sold by roughly 32 cents, an increase of 7.8% relative to the mean value for non-inventory costs per dollar sold across the sample.

Panels C and D focus on wholesale establishments. For these models, a one standard deviation increase in $\tilde{G}_{Ret,Emp}$ for neighborhood retail activity has a significant effect on wholesale non-inventory costs per dollar of sales. The point estimates are 13.2 cents in Panel C and 20.9 cents in Panel D. These

[33] See https://www.bls.gov/cew/data.htm for BLS estimates and The Retail Sector in New York City: Recent Trends and the Impact of COVID-19 - December 2020 (state.ny.us) for discussion by the New York State comptroller.

estimates suggest that wholesalers also benefit from enhanced eyes on the street associated with retail spatial concentration, but to a lesser degree than retail establishments. This is as expected.[34]

5. Conclusions

Busy city streets are often thought to deter crime by amplifying the effect of "eyes on the street," offsetting the potential for criminals to hide in a crowd (e.g. Jacobs, 1961; Jarrell and Howsen, 1990; Harries, 2006; Browning and Jackson, 2013; Chang and Jacobson, 2017; Carr and Doleac, 2018; McMillen et al., 2019; Tillyer and Walter, 2019). This paper provides new support for this idea, focusing on the retail sector. U.S. retailers lose over 2% of sales to property crime each year, a substantial amount relative to profit margins that average roughly 3%.[35] Local authorities also devote considerable resources to patrolling retail districts. We show that concentrating retail establishments at the street level has potential to reduce these costs.

Using data for New York City, findings indicate that for a shift from the 25th to the 75th percentile neighborhood based on retail spatial concentration, property crime decreases by 9.4%, police stops are reduced by 12.1%, and retail rent increases by 9.6%. These estimates are robust to alternate model designs, and under plausible conditions, likely understate the effect of crime deterrence associated with spatial concentration. Additional findings suggest that these effects arise from a combination of crowding and visibility that are enhanced by spatial concentration of retail activity.

Together the various models and estimates in our paper suggest that block-level spatial concentration of retail activity enhances eyes-on-the-street, and much more so than concentration of other industries. Our estimates are also large enough to be important. Local government and the private sector can reduce the cost of crime by encouraging retailers to concentrate spatially within their neighborhoods.

---

[34] Other coefficients in the retail and wholesale models in Table 5 are in line with priors. This includes establishment age fixed effects. These display a strong monotonic pattern in which older companies have smaller non-inventory costs per dollar sold. This is consistent with priors that older companies that have survived a competitive weeding out process are more productive and enjoy lower cost-to-sale ratios than younger establishments.

[35] See discussion by the Small Business Resource Center at https://sbrc.employers.com/retail/whats-a-good-profit-margin-for-retailers/ and the associated report by Deloitte Inc. (2018).

References

Becker, G. (1968). Crime and Punishment: An economic approach. *Journal of Political Economy*, 76 (2), 169–217.

Billings, S.B., Deming, D.J. and Ross, S.L. (2019). Partners in Crime. *American Economic Journal: Applied Economics*, 11(1), 126-150.

BLS -U.S. Bureau of Labor and Statistics-.(2018). The Quarterly Census of Employment and Wages. Accessed on April 27th, 2020. https://www.bls.gov/cew/data.htm.

Bowes, D. R. (2007). A Two-Stage Model of the Simultaneous Relationship Between Retail Development and Crime. *Economic Development Quarterly*, 21(1), 79–90.

Brooks, L. (2008). Volunteering to Be Taxed: Business Improvement Districts and the Extra-Governmental Provision of Public Safety. *Journal of Public Economics*, 92(1-2), 388-406.

Browning, C.R.; Byron, R.; Calder, C.A.; Krivo, L.J.; Kwan, M.P.; Lee, J.Y. and Peterson, R.D. (2010). Commercial Density, Residential Concentration, and Crime: Land Use Patterns and Violence in Neighborhood Context. *Journal of Research in Crime and Delinquency*, 47(3), 329–357.

Browning, C.R. and Jackson, A.L. (2013). The Social Ecology of Public Space: Active Streets and Violent Crime in Urban Neighborhoods. *Criminology*, 51, 1009–1043.

Carr, J.B and Doleac, J.L. (2018). Keep the Kids Inside? Juvenile Curfews and Urban Gun Violence. *The Review of Economics and Statistics*, 100(4), 609–618.

Chang, T.Y. and Jacobson, M. (2017). Going to pot? The impact of dispensary closures on crime. *Journal of Urban Economics*, 100, 120–136.

Clapp, J.M.; Ross, S.R. and Zhou, T. (2019). Retail Agglomeration and Competition Externalities: Evidence from Openings and Closings of Multiline Department Stores in the U.S. *Journal of Business & Economic Statistics*, 37(1), 81–96.

Deloitte Inc. (2018). "Global Powers of Retailing 2018 Transformative change, reinvigorated commerce", https://www2.deloitte.com/content/dam/Deloitte/at/Documents/about-deloitte/global-powers-of-retailing-2018.pdf .

Ellen, I.G.; Lacoe, J. and Sharygin, C.A. (2013). Do foreclosures cause crime? *Journal of Urban Economics*, 74, 59–70.

Evans, D. N.; Maragh, C. L. and Porter, J. R. (2014). What do We Know About NYC's Stop and Frisk Program? A Spatial and Statistical Analysis. *Advances in Social Sciences Research Journal*, 1(2), 129–144.

Faggio, G. (2021). The impact of Business Improvement Districts on crime. Working Paper.

Ferrandino, J.A. (2018). The effectiveness and equity of NYPD stop and frisk policy, 2003–2014. *Journal of Crime and Justice*, 41(2), 119–135.

Freedman, M. and E. Owens (2016). Your Friends and Neighbors: Localized Economic Development, Inequality, and Criminal Behavior. *Review of Economics and Statistics*, 98(2), 233–253.

Gelman, A.; Fagan, J. and Kiss, A. (2007). An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias. *Journal of the American Statistical Association*, 102(479), 813–823,

Getis, A. and Ord, J.K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, 2 (3), 189–206.

Gonzalez, R. and Komisarow, S. (2020). Community monitoring and crime: Evidence from Chicago's Safe Passage Program. *Journal of Public Economics*, 191, 104250.

Gould, E.D., Pashigian, B.P. and Prendergast, C.G. (2005). Contracts, Externalities, and Incentives in Shopping Malls. *The Review of Economics and Statistics*, 87(3), 411–422.

Greenbaum, R. T., and Tita, G. E. (2004). The Impact of Violence Surges on Neighbourhood Business Activity. *Urban Studies*, 41(13), 2495–2514.

Groff, E. R., and Lockwood, B. (2014). Criminogenic Facilities and Crime across Street Segments in Philadelphia: Uncovering Evidence about the Spatial Extent of Facility Influence. *Journal of Research in Crime and Delinquency*, 51(3), 277–314.

Hakim, S. and Shachmurove, Y. (1996). Spatial and Temporal Patterns of Commercial Burglaries: The Evidence Examined. *The American Journal of Economics and Sociology*, 55(4), 443–456.

Hanink, P. (2013). Don't Trust the Police: Stop Question Frisk, Compstat, and the High Cost of Statistical over-Reliance in the NYPD. *Journal of the Institute of Justice and International Studies*, 13, 99–114.

Harries, K. (2006). Property Crimes and violence in United States: An Analysis of the Influence of Population Density. *International Journal of Criminal Justice Sciences*, 1(2), 24-34.

Hipp, J.R. (2016). General Theory of Spatial Crime Patterns. *Criminology*, 54(4), 653–679.

Hoyt, L. M. (2005). Do Business Improvement District Organizations Make a Difference?: Crime in and Around Commercial Areas in Philadelphia. *Journal of Planning Education and Research*, 25(2), 185–199.

Jacobs, J. (1961). The Death and Life of Great American Cities. New York: Vintage Books.

Jarrell, S. and Howsen, R.M. (1990). Transient Crowding and Crime: The more "Strangers" in an Area, the More Crime Except for Murder, Assault and Rape. *American Journal of Economics and Sociology, Inc.* 49(4), 483-494.

Johansen, B.O. and Nilssen, T. (2016). The Economics of Retailing Formats. *Journal of Industrial Economics*, 64(1), 109–134.

Koster, H.R.A.; van Ommeren, J. and Rietveld, P. (2014), Agglomeration Economies and Productivity: A Structural Estimation Approach Using Commercial Rents. *Economica*, 81(321), 63–85.

Koster, H.R.A.; Pasidis, I. and van Ommeren, J. (2019). Shopping externalities and retail concentration: Evidence from dutch shopping streets. *Journal or Urban Economics*, 114, 103194.

Lee, M.R. and Alshalan, A. (2005). Geographic Variation in Property Crime Rates: A Test of Opportunity Theory. *Journal of Crime and Justice*, 28(2), 101–127.

Lens, M.C. and Meltzer, R. (2016), Is Crime Bad for Business? Crime and Commercial Property Values In New York City. *Journal of Regional Science*, 56(3), 442–470.

Linden, L. and Rockoff, J.E. (2008). Estimates of the Impact of Crime Risk on Property Values from Megan's Laws. *American Economic Review*, 98(3), 1103–1127.

MacDonald, J. and Braga, A.A. (2019). Did Post-*Floyd et al.* Reforms Reduce Racial Disparities in NYPD Stop, Question, and Frisk Practices? An Exploratory Analysis Using External and Internal Benchmarks, *Justice Quarterly*, 36 (5), 954–983.

MacDonald, J.; Fagan, J. and Geller, A. (2016). The Effects of Local Police Surges on Crime and Arrests in New York City. *PLoS ONE*, 11 (6): e0157223.

McMillen, D.; Sarmiento-Barbieri, I. and Singh, R. (2019). Do more eyes on the street reduce Crime? Evidence from Chicago's safe passage program. *Journal of Urban Economics*, 110, 1–25.

NRF-National Retail Federation-. (2019). National Retail Security Survey. Accessed on March 29th, 2021. https://nrf.com/research/national-retail-security-survey-2019

NYCC-New York City Council (2017). Report to the Committee on Finance and the Committee on Public Safety on the Fiscal 2018 Executive Budget for New York Police Department. Accessed on March 29th, 2021. http://council.nyc.gov/budget/wp-content/uploads/sites/54/2017/03/056-NYPD-exec-1.pdf

Ord, J.K. and Getis, A. (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*, 27 (4), 286–306.

Owens, E. (2019). Policy Essay: An Economic Approach to De-Policing. *Criminology and Public Policy*, 18(1), 77–80.

Owens, E. (2020). The Economics of Policing. In D. Marcotte and K. Zimmermann (Eds.), *Economics of Risky Behavior*. Springer.

Pashigian, B. and Gould, E. (1998). Internalizing Externalities: The Pricing of Space in Shopping Malls. *The Journal of Law and Economics*, 41(1), 115–142.

Pope, J. (2008). Fear of Crime and Housing Prices: Household Reactions to Sex Offender Registries. *Journal of Urban Economics*, 64, 601–614.

Ridgeway, G. (2007). Analysis of Racial Disparities in the New York Police Department's Stop, Question, and Frisk Practices. *RAND Corporation*, TR-534-NYCPF, 2007. As of February 24, 2021: https://www.rand.org/pubs/technical_reports/TR534.html

Robinson, P.M. (1988). Root-n-consistent semiparametric regression. Econometrica 56, 931–954.

Rosenfeld, R. and Fornango, R. (2017). The Relationship Between Crime and Stop, Question, and Frisk Rates in New York City Neighborhoods. *Justice Quarterly*, 34(6), 931–951.

Rosenthal, S.S. and Ross, A. (2010). Violent crime, entrepreneurship, and cities. *Journal of Urban Economics*, 67(1), 135–149.

Rosenthal, S.S. and Strange, W.C. (2020). How Close Is Close? The Spatial Reach of Agglomeration Economies. *Journal of Economic Perspectives*, 34 (3): 27–49.

Rosenthal, S.S.; Strange, W.C. and Urrego, J.A. (2021). Are city centers losing their appeal? Commercial real estate, urban spatial structure, and COVID-19. *Journal of Urban Economics: Insights*, forthcoming.

Sivitanidou, R. (1995). Urban Spatial Variations in Office-Commercial Rents: The Role of Spatial Amenities and Commercial Zoning. *Journal of Urban Economics*, 38(1), 23–49.

Stucky, T.D. and Ottensmann, J.R. (2009). Land Use and Violent Crime. *Criminology*, 47(4), 1223–1264.

Tillyer, M. S., and Walter, R. J. (2019). Busy Businesses and Busy Contexts: The Distribution and Sources of Crime at Commercial Properties. *Journal of Research in Crime and Delinquency*, 56(6), 816–850.

Urrego, J.A. (2023). Perceptions of Police Bias and the Location of Minority-Owned Establishments. Working paper.

Verardi, V. and Debarsy, N. (2012). Robinson's square root of N consistent semiparametric regression estimator in Stata. *The Stata Journal*, 12(4), 726–735.

Weisburd, D.; Wooditch, A.; Weisburd, S. and Yang, S.M. (2016). Do Stop, Question, and Frisk Practices Deter Crime? Evidence at microunits of space and time. *Criminology & Public Policy*, 15 (1), 31–56.

Weterings, A. (2014). What Makes Firms Leave the Neighbourhood? *Urban Studies*, 51(8), 1613–1633.

Wooditch, A. and Weisburd, D. (2016). Using Space–Time Analysis to Evaluate Criminal Justice Programs: An Application to Stop-Question-Frisk Practices. *Journal of Quantitative Criminology*, 32 (2), 191–213.

Figure 1a: Retail Employment in Included Neighborhoods[a]

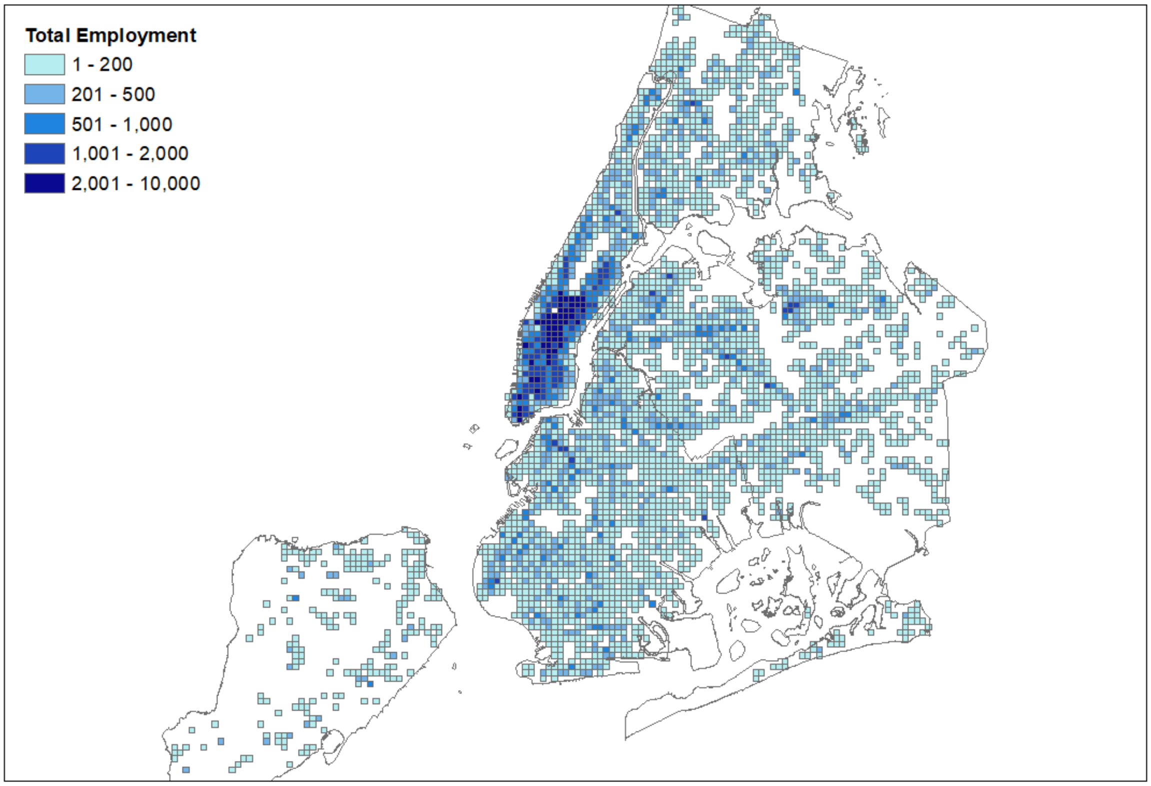


Figure 1b: Spatial Concentration of Retail Employment ($\tilde{G}$) in Included Neighborhoods[a]

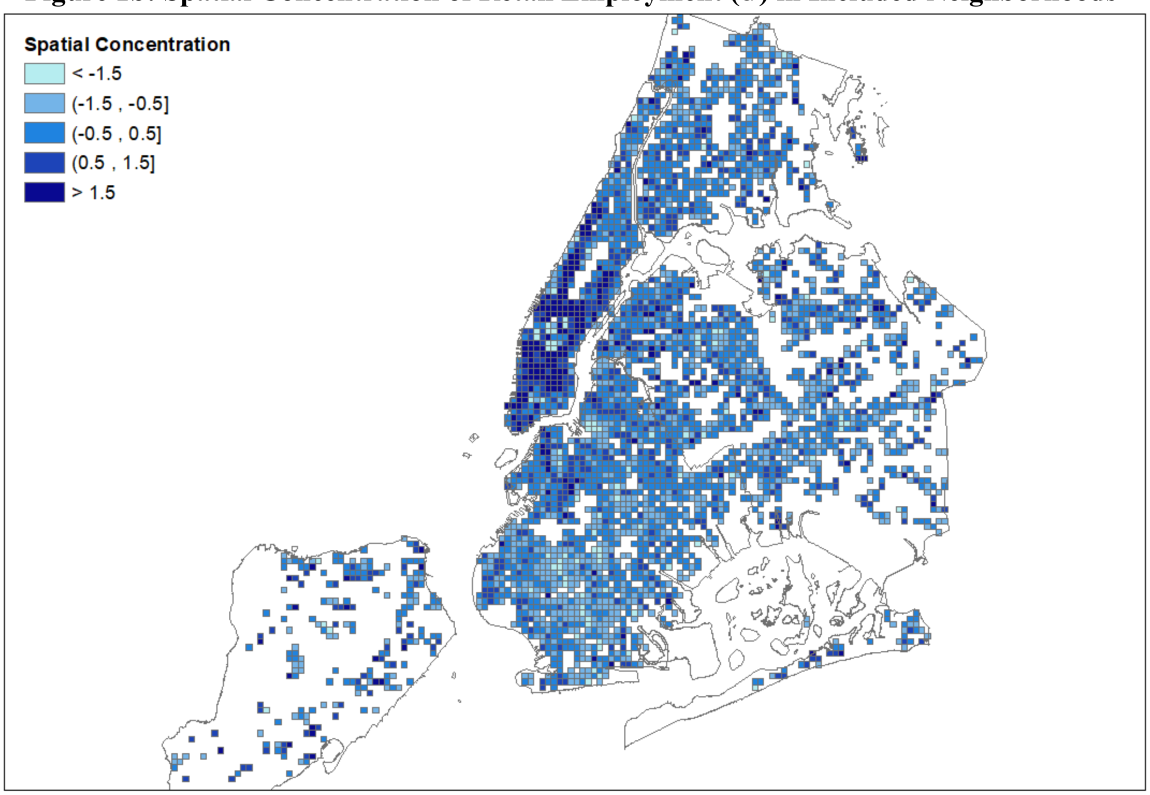
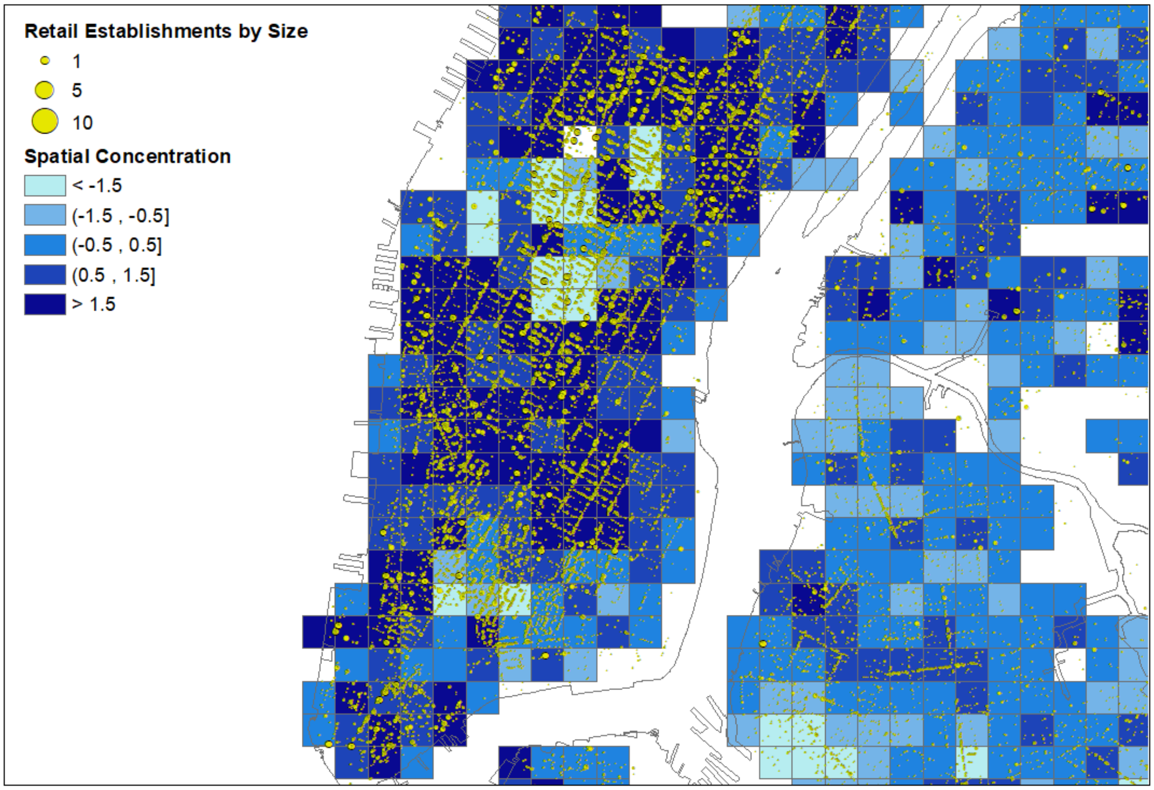
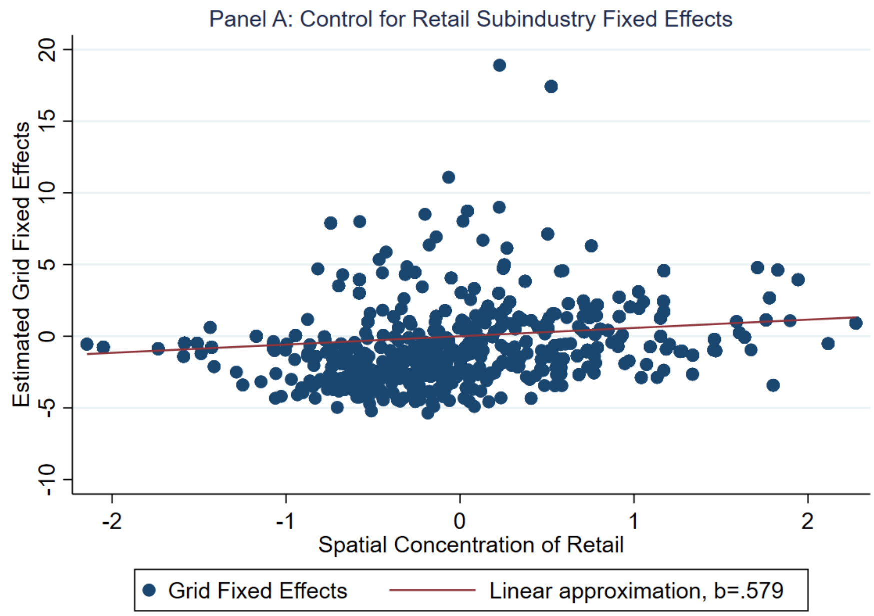


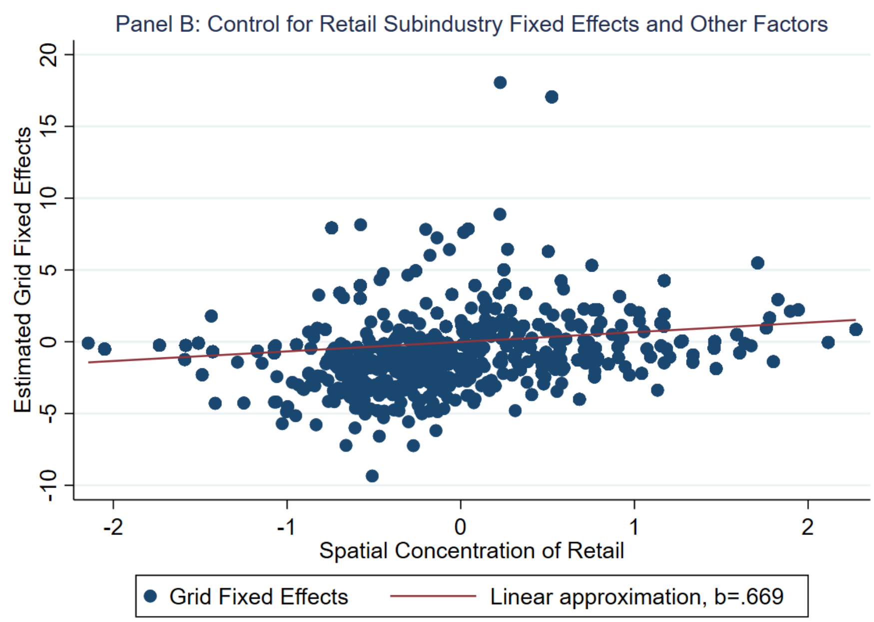
Figure 1c: Retail Employment and Spatial Concentration South of Central Park[a]



[a] Figure 1a and 1b plot the 3,506 grid squares included in the primary estimating sample extending over all 5 Boroughs of New York City. In Figures 1b and 1c, spatial concentration of retail employment, $\tilde{G}_{Retail,Emp}$, is calculated as described in expression (3.3a). Figure 1c covers the area south of Central Park to the Financial District in Manhattan. Total retail employment in Figure 1a includes employment from establishments belonging to single-site and multi-site firms. Retail spatial concentration in Figures 1b and 1c is based on employment at just single-site firms for reasons described in the text, as is the scatter plot of individual retail establishments in Figure 1c.

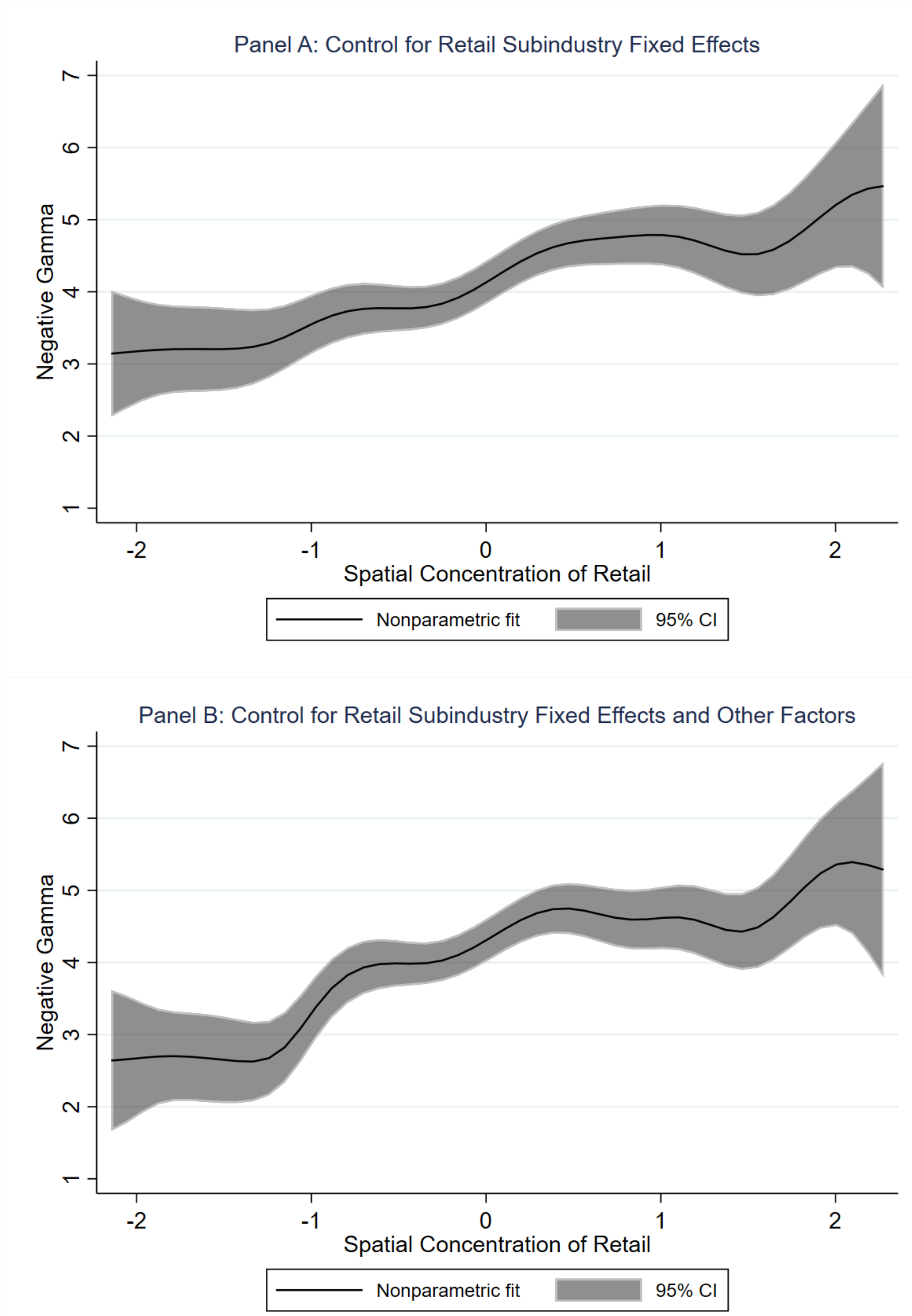
Figure 2: Neighborhood square fixed effect estimates of - $\gamma(\tilde{G})$ plotted against $\tilde{G}$[a]





[a] Panel A plots the estimated grid square fixed effects from expression (2.7). Panel B does the same but includes controls for other neighborhood and establishment attributes as in (2.8) including spatial concentration of other industries, visits to POI, share of residential units, number of trees and dummies for different age categories of retail establishments.

Figure 3: Partial linear model estimates of the effect of $\tilde{G}$ on - $\gamma(\tilde{G})$[a]



[a] Panel A plots the estimated $-\gamma(\tilde{G})$ function from a partial linear model where non-inventory cost per dollar of sales depends, in addition to γ, on subindustry fixed effects. Panel B adds other controls linearly to the specification in Panel A. Those are spatial concentration of other industries, visits to POI, share of residential units, number of trees and fixed effect for different age categories of retail establishments.

Table 1a: Attributes of Included and Omitted Neighborhood Grid Squares

Panel A: Grid Squares included in our Estimating Sample

Variable	Obs.	Mean	Std. Dev.	p10	p50	p90
Total Number of Establishments	3,506	238.61	462.03	49	134	399
Total Employment	3,506	1,542	5,023	159	563	2,198
Employment: Retail	3,506	229	530	15	96	466
Employment: Finance	3,506	213	1,501	6	30	174
Employment: Manufacturing	3,506	101	577	2	10	123
Employment: Services	3,506	725	2,240	62	232	1,146
Share of Residential Units Within Buildings	3,506	0.87	0.20	0.696	0.937	0.992

Panel B: Grid Squares not included in our Estimating Sample

Variable	Obs.	Mean	Std. Dev.	p10	p50	p90
Total Number of Establishments	2,727	40.34	43.42	5	33	82
Total Employment	2,727	214	526	19	107	444
Employment: Retail	2,727	31	99	0	9	68
Employment: Finance	2,727	11	64	0	4	22
Employment: Manufacturing	2,727	4	28	0	0	5
Employment: Services	2,727	117	379	5	46	231
Share of Residential Units Within Buildings	2,708	0.86	0.30	0.210	0.988	1.000

Table 1b: Summary Statistics[a]

	Obs.	Mean	Std. Dev.	p25	p50	p75
Panel A: At the grid square level[b]						
Property crime	3,506	29.06	54.78	6	14	31
Grand Larceny & Burglary	3,506	9.76	16.37	2	5	11
Petit Larceny	3,506	18.40	40.45	3	8	18
Theft of Services & Fraud	3,506	0.90	2.13	0	0	1
Robbery	3,506	2.57	3.48	0	1	4
Auto Theft	3,506	0.95	1.18	0	1	1
Police Stops	3,506	2.08	4.37	0	0	2
At Least 1 Police Stop (Stops > 0)	3,506	0.49	0.50	0	0	1
Share of Employment						
Retail	3,506	0.19	0.12	0.09	0.17	0.27
Finance	3,506	0.07	0.07	0.03	0.06	0.09
Manufacture	3,506	0.05	0.08	0.01	0.02	0.05
Services	3,506	0.46	0.18	0.34	0.46	0.58
Spatial Concentration[d]						
$\tilde{G}_{Retail,Emp}$	3,506	0.00	1.00	-0.63	-0.11	0.47
$\tilde{G}_{Finance,Emp}$	3,506	0.00	1.00	-0.54	-0.21	0.26
$\tilde{G}_{Manf,Emp}$	3,506	0.00	1.00	-0.67	-0.24	0.40
$\tilde{G}_{Service,Emp}$	3,506	0.00	1.00	-0.52	-0.19	0.23
Number of Trees	3,506	119.25	55.53	80	118	157
Average Age Buildings	3,506	80.26	17.04	70.48	82.44	91.83
Average Building Assessment	3,506	2,116,857	15,500,000	75,979	177,548	665,259
Overlaps Multiple Police Precinct	3,506	0.24	0.43	0	0	0
Share of Residential Units Within Bldgs.	3,506	0.87	0.20	0.87	0.94	0.97
Neighborhood Sales per Worker	3,506	58,934	60,118	40,379	51,627	65,304
Average Monthly Visitors POI	3,506	179.91	118.88	104.92	148.85	215.43

	Obs.	Mean	Std. Dev.	p25	p50	p75
Panel B: At the establishment level[c]						
Spatial Concentration of Retail[d]	1,596	0.01	0.98	-0.57	0.02	0.65
Retailer Rent per sq foot (psf) per month	1,596	178.02	211.27	61.38	110.98	219.75
Retailer Employment	1,596	7.21	11.49	2	3	8
Retailer Sales	1,596	471,918	1,438,367	89,625	150,000	289,590
Retailer Cost of space / Sales	1,596	2.95	4.03	0.55	1.38	3.59
Retailer Non-inventory cost / Sales[e]	1,596	4.15	4.29	1.45	2.73	5.27
Wholesaler Cost of space / Sales	538	0.66	1.06	0.06	0.21	0.74
Wholesaler Non-inventory cost / Sales[e]	538	1.44	1.41	0.43	0.97	1.89

[a] Crime and police stops data are from the New York Police Department. Crimes lasting more than one day and all crimes that take place in a transportation system (e.g., on the subway) are dropped. Police stops prompted by 911 calls and those for ongoing investigations are dropped. Employment and sales are from Dun & Bradstreet while rent and space leased are from CompStak.
[b] The unit of analysis is a grid cell of 0.2 square miles across the New York City area.
[c] Panel B includes only establishments matched in the CompStak and Dun and Bradstreet files.
[d] Spatial concentration is calculated as described for $\tilde{G}$ in the text.
[e] Non-inventory cost refers to the sum of cost of space and labor cost. Estimated annual earnings from BLS at the NAICS 6 digits code are used to calculate labor cost.

Table 2: Negative Binomial Marginal Effects for Property Crime and Police Stops in 0.2 Mile Grid Squares[a]

	Property Crime[b]					Police Stops[c]				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Include Residential Neigh	Exclude Residential Neigh	Exclude Multi-Site Firms for $\tilde{G}_{Emp}$	Include Multi-Site Firms for $\tilde{G}_{Emp}$	9 Square Neigh Design[c]	Include Residential Neigh	Exclude Residential Neigh	Exclude Multi-Site Firms for $\tilde{G}_{Emp}$	Include Multi-Site Firms for $\tilde{G}_{Emp}$	9 Square Neigh Design[c]
Log total employment	0.722***	0.700***	0.554***	0.547***	0.665***	0.539***	0.394***	0.349***	0.341***	0.378***
	(0.024)	(0.017)	(0.022)	(0.023)	(0.014)	(0.031)	(0.037)	(0.056)	(0.056)	(0.036)
Share of Emp: Retail	2.557***	2.651***	2.307***	2.247***	2.879***	2.002***	2.411***	2.017***	2.319***	1.643***
	(0.202)	(0.135)	(0.165)	(0.166)	(0.163)	(0.254)	(0.279)	(0.359)	(0.362)	(0.388)
Share of Emp: Finance	-	-	-0.451**	-0.577**	-1.002***	-	-	-0.752	-0.217	-3.230***
	-	-	(0.217)	(0.226)	(0.204)	-	-	(0.608)	(0.625)	(0.556)
Share of Emp: Manufacture	-	-	-0.226	-0.108	0.412	-	-	0.027	0.325	-0.650
	-	-	(0.247)	(0.260)	(0.321)	-	-	(0.536)	(0.546)	(0.637)
Share of Emp: Services	-	-	-0.025	-0.044	0.363***	-	-	0.055	-0.183	-0.101
	-	-	(0.129)	(0.129)	(0.114)	-	-	(0.318)	(0.319)	(0.269)
Log monthly visits to POI	-	-	0.516***	0.526***	0.611***	-	-	0.894***	0.900***	1.284***
	-	-	(0.030)	(0.030)	(0.023)	-	-	(0.069)	(0.071)	(0.055)
Retail ($\tilde{G}_{Ret,Emp}$)[d]	-0.097***	-0.061***	-0.085***	-0.056***	-1.058***	-0.150***	-0.087**	-0.110***	-0.223***	-2.235***
	(0.016)	(0.018)	(0.015)	(0.016)	(0.125)	(0.036)	(0.040)	(0.036)	(0.044)	(0.322)
Finance ($\tilde{G}_{Fin,Emp}$)[d]	-	-	0.005	0.017	-0.179	-	-	0.042	-0.048	0.693***
	-	-	(0.016)	(0.018)	(0.112)	-	-	(0.040)	(0.039)	(0.234)
Manufacture ($\tilde{G}_{Manf,Emp}$)[d]	-	-	0.007	-0.012	0.036	-	-	-0.020	-0.071	-0.357***
	-	-	(0.018)	(0.019)	(0.062)	-	-	(0.041)	(0.045)	(0.132)
Services ($\tilde{G}_{Serv,Emp}$)[d]	-	-	-0.029	-0.012	-0.647***	-	-	-0.137***	0.057	-1.120***
	-	-	(0.019)	(0.020)	(0.145)	-	-	(0.046)	(0.046)	(0.248)
Other Neighborhood Controls[e]	No	No	6	6	6	No	No	6	6	6
Overdispersion Poisson	0.69	0.58	0.42	0.421	0.217	3.208	2.59	2.03	2.004	1.115
Observations	5,461	3,506	3,506	3,546	5,122	5,461	3,506	3,506	3,546	5,122

[a] Marginal effects based on the data means are reported. Significance is denoted as: * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors in parentheses.
[b] Property crime includes 2018 incidents of grand larceny, burglary, petit larceny, theft of services, and fraud.
[c] Police stops include discretionary stops pooled from 2016-2018.
[d] Spatial concentration is calculated as described for $\tilde{G}$ in expression (3.3a).
[e] Neighborhood controls include log number of trees, average building age, log of average building assessment, whether the grid square overlaps with more than one police precinct, share of residential units within buildings, and log of neighborhood sales per worker.

Table 3: Number of Crimes for Alternate Time Periods by Type of Crime[a,b]

	All Hours and Periods		Daytime[c]		Nighttime[c]		COVID-19 Lockdown[d]	
	% Retail Employment	$\tilde{G}_{Ret,Emp}$	% Retail Employment	$\tilde{G}_{Ret,Emp}$	% Retail Employment	$\tilde{G}_{Ret,Emp}$	% Retail Employment	$\tilde{G}_{Ret,Emp}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Property crime	2.307***	-0.085***	2.650***	-0.104***	1.685***	-0.004	2.647***	-0.034
	(0.165)	(0.015)	(0.182)	(0.018)	(0.198)	(0.018)	(0.339)	(0.032)
Grand Larceny & Burglary	1.792***	-0.053***	2.034***	-0.072***	1.766***	0.003	2.493***	-0.070*
	(0.159)	(0.014)	(0.179)	(0.016)	(0.223)	(0.021)	(0.424)	(0.040)
Petit Larceny	2.650***	-0.104***	3.050***	-0.121***	1.733***	-0.008	2.810***	-0.009
	(0.189)	(0.018)	(0.208)	(0.021)	(0.238)	(0.022)	(0.418)	(0.040)
Theft of Services & Fraud	1.819***	-0.087***	2.204***	-0.125***	1.298***	-0.007	2.641*	0.001
	(0.319)	(0.030)	(0.404)	(0.042)	(0.459)	(0.039)	(1.474)	(0.279)
Robbery	2.330***	-0.144***	2.217***	-0.175***	2.265***	-0.142***	3.352***	-0.177***
	(0.220)	(0.021)	(0.254)	(0.027)	(0.265)	(0.028)	(0.669)	(0.058)
Auto Theft	0.833***	-0.057**	0.586*	-0.042	1.152***	-0.086***	1.018	0.008
	(0.212)	(0.023)	(0.333)	(0.034)	(0.320)	(0.033)	(0.854)	(0.087)

[a] Marginal effects based on the data means are reported. Significance is denoted as: p<0.1, ** p<0.05, *** p<0.01. Robust standard errors in parenthesis. Spatial concentration is calculated as described for $\tilde{G}$ in expression (3.3a).
[b] Every pair of coefficients, share of retail employment and spatial concentration of retail, is obtained from the preferred specification in Table 2 column (3). Note that the coefficients for property crime during all hours are identical to those in Table 2 column (3). The sample for all models is 3,506 grid cells (0.2 square miles) in the NYC area. Crime data refers to incidents in 2018.
[c] Daytime hours include crimes between 10 am and 6 pm. Nighttime hours include crimes between 10 pm and 5 am.
[d] COVID-19 lockdown refers to the first two weeks of the NYC lockdown, March 22nd to April 5th of 2020.

Table 4: Alternate Measures of Spatial Concentration[a]

Panel A: Correlation between Alternate Measures of Retail Spatial Concentration

	$\tilde{G}_{Ret,Emp}$	$\tilde{G}_{Ret,Stores}$	$\tilde{G}_{Ret,Sales}$
$\tilde{G}_{Ret,Emp}$	1.0	-	-
$\tilde{G}_{Ret,Stores}$	0.15	1.0	-
$\tilde{G}_{Ret,Sales}$	0.54	0.14	1.0

Panel B: Number of Property Crimes Controlling for Different Measures of Spatial Concentration[b]

	(1)	(2)	(3)	(4)
Regression Sample	% Retail Employment	$\tilde{G}_{Ret,Emp}$ [c]	$\tilde{G}_{Ret,Stores}$ [c]	$\tilde{G}_{Ret,Sales}$ [c]
All Hours/Periods	2.646***	-0.084***	-0.122***	0.002
	(0.178)	(0.016)	(0.021)	(0.016)
Daytime	2.974***	-0.101***	-0.120***	-0.001
	(0.198)	(0.019)	(0.025)	(0.018)
Nighttime	2.109***	-0.021	-0.141***	0.028
	(0.210)	(0.020)	(0.024)	(0.020)
COVID-19 Lockdown	3.096***	-0.078**	-0.142***	0.052
	(0.355)	(0.036)	(0.048)	(0.036)

[a] Marginal effects based on the data means are reported. Significance is denoted as: * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors in parenthesis. The sample for all models is 3,506 grid cells (0.2 square miles) in the NYC area.

[b] Property crime includes 2018 incidents of grand larceny, burglary, petit larceny, theft of services, and fraud.

[c] Spatial concentration of retail employment, $\tilde{G}_{\text{Ret,Emp,}}$ is calculated as described for $\tilde{G}$ in expression (3.3). Similar structure is applied to the spatial concentration of retail sales, $\tilde{G}_{Ret,Sales}$, where establishment's sales is used instead of employment in (3.3a). Spatial concentration of storefronts, $\tilde{G}_{\text{Ret,Stores}}=\sum_e \omega_{ie}(d_{ie})\ /n_i(d_{ie})$, where n_i is the number of nearby establishments within a given distance, d_{ie}, and ω_{ie} is defined as in (3.2).

Table 5: Cost of Inventory Lost to Crime Per Dollar Sold[a]

	(1)	(2)	(3)	(4)	(5)
Panel A: Retailer cost of space/$ sold					
Number of Workers/Sales	71,574***	70,893***	71,129***	71,828***	37,983***
	(6,196)	(6,116)	(6.116)	(6,432)	(7,180)
Spatial Concentration of Retail: $\tilde{G}_{Ret,Emp}$	0.373***	0.413***	0.436***	0.384***	0.353***
	(0.106)	(0.116)	(0.115)	(0.123)	(0.120)
Observations	1,596	1,596	1,596	1,596	1,596
R2	0.47	0.48	0.48	0.50	0.52
Panel B: Retailer non-inventory cost/$ sold					
Spatial Concentration of Retail: $\tilde{G}_{Ret,Emp}$	0.457***	0.517***	0.533***	0.377***	0.324***
	(0.119)	(0.132)	(0.132)	(0.141)	(0.125)
Observations	1,596	1,596	1,596	1,596	1,596
R2	0.538	0.544	0.544	0.567	0.63
Panel C: Wholesaler cost of space/$ sold					
Number of Workers/Sales	84,814***	82,089***	81,671***	81,802***	70,405***
	(8,355)	(8,707)	(8,864)	(9,410)	(9,734)
Spatial Concentration of Retail: $\tilde{G}_{Ret,Emp}$	0.168***	0.185***	0.177***	0.162***	0.132**
	(0.053)	(0.056)	(0.054)	(0.058)	(0.058)
Observations	538	538	538	538	538
R2	0.471	0.49	0.49	0.529	0.549
Panel D: Wholesaler non-inventory cost/$ sold					
Spatial Concentration of Retail $\tilde{G}_{Ret,Emp}$	0.247***	0.288***	0.253***	0.244***	0.209**
	(0.073)	(0.078)	(0.074)	(0.080)	(0.083)
Observations	538	538	538	538	538
R2	0.579	0.59	0.593	0.626	0.685
Neigh Controls (Table 2) and Corner location	Yes	Yes	Yes	Yes	Yes
SIC 2 Fixed Effects[b]	Yes	Yes	Yes	Yes	Yes
Lease Execution Year Fixed Effects[b]	Yes	Yes	Yes	Yes	Yes
Spatial Concentration other industries	-	Yes	Yes	Yes	Yes
POI Visits	-	-	Yes	Yes	Yes
Police Precinct Fixed Effects[c]	-	-	-	Yes	Yes
Establishment Age Categories[d]	-	-	-	-	Yes

[a] Significance is denoted as: * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors are in parenthesis. Rent data comes from CompStak and employment and sales from Dun & Bradstreet. Establishments located above the 25th floor are dropped as are those in the top 1% of the distribution of sales, employment, space leased, and employment divided by sales. Non-inventory cost is the sum of expenditures on space and labor. Establishments with non-inventory cost in the top 5% are dropped.
[b] Models in Panels A and B include 24 lease transaction year fixed effects and 8 SIC2 fixed effects. Panels C and D include 19 lease transaction year fixed effects and 2 SIC2 fixed effects.
[c] In column (4), 69 police precinct fixed effects are present for the retail regressions and 34 for the wholesale regressions.
[d] Column (5) includes fixed effects for establishment age categories: less than 2 years, 2 to 5 years, 6 to 10 years, 10 to 25 years, and more than 25 years in business.

Appendix A: Efficient Levels of Protection and Crime

This appendix solves for the efficient levels of protection and crime. We begin by solving for the cost of providing P_j units of protection to each store, $C_{P,j}$. This is obtained by substituting (2.4a) and (2.4b) into (2.3) in the text (the optimal levels of investment in protection and expenditure on protection, respectively). Grouping parameters into a positive constant κ and solving,[36]

$$C_{P,j} = \kappa \left[\frac{P_j}{G_j}\right]^{\frac{1}{\alpha+\beta}} \left[\frac{1}{N_j}\right]^{\frac{\alpha}{\alpha+\beta}} \quad . \tag{A.1}$$

We also express the effect of a given level of protection on inventory that is sold from each store as,

$$I_j^{sold} = I_j(1 - P_j^{-\lambda}) \text{ where } \lambda > 0 \text{ and } P_j > 1 \; . \tag{A.2}$$

In this expression, deterrence increases with λ as criminal skill at evading protection measures diminishes, or equivalently, as penalties that discourage criminal activity become increasingly severe.

Subtracting (A.1) from (A.2) yields a measure of the net gain to each store from its expenditure on private and public protection against crime. Taking first order conditions and rearranging,[37]

$$P_j^* = \lambda^{\frac{\alpha+\beta}{1+\lambda(\alpha+\beta)}} I_j^{\frac{\alpha+\beta}{1+\lambda(\alpha+\beta)}} G_j^{\frac{1}{1+\lambda(\alpha+\beta)}} N_j^{\frac{\alpha}{1+\lambda(\alpha+\beta)}} \left(\frac{\alpha+\beta}{\kappa}\right)^{\frac{\alpha+\beta}{1+\lambda(\alpha+\beta)}} . \tag{A.3}$$

Inventory lost to crime is then obtained by substituting (A.3) into (A.2) and using the identity in (2.1),

$$I_j^{stolen} = \left(\frac{1}{\lambda}\right)^{\frac{\lambda(\alpha+\beta)}{1+\lambda(\alpha+\beta)}} I_j^{\frac{1}{1+\lambda(\alpha+\beta)}} \left(\frac{1}{G_j}\right)^{\frac{\lambda}{1+\lambda(\alpha+\beta)}} N_j^{-\frac{\lambda\alpha}{1+\lambda(\alpha+\beta)}} \left(\frac{\alpha+\beta}{\kappa}\right)^{-\frac{\lambda(\alpha+\beta)}{1+\lambda(\alpha+\beta)}} \tag{A.4}$$

From (A.3) and (A.4), higher G_j increases protection against crime while reducing the amount of inventory stolen.[38] This follows from the core modeling assumption that spatial concentration enhances the effect of eyes on the street.

---

[36] $\kappa = \left(\frac{\alpha}{\beta}\right)^{-\frac{\alpha}{\alpha+\beta}} + \left(\frac{\alpha}{\beta}\right)^{\frac{\beta}{\alpha+\beta}}$

[37] The first order condition requires that $\lambda I_j P_j^{-(\lambda+1)} - \frac{\kappa}{\alpha+\beta} G_j^{-\frac{1}{\alpha+\beta}} N_j^{-\frac{\alpha}{\alpha+\beta}} P_j^{\frac{1}{\alpha+\beta}-1} = 0$.

[38] Notice also that as λ shrinks to zero in (A.4), I_j^{stolen} converges to I_j and all inventory is stolen regardless of P. This however would cause all establishments to exit the market. More generally, shutdown decisions could be made explicit by deriving a maximum share of inventory lost to crime, beyond which profit in the following section becomes negative. Because this would not affect our primary results, we adopt the simpler specification above.

Appendix B: Data Sources, Access, and Variable Construction

B.1 Matching D&B and CompStak establishment level data

For the rent analysis in Table 5 we used establishment level matched records from D&B and CompStak. Our match routine took advantage of street addresses and establishment names which are reported in both files and utilize two similarity indexes. One uses n-gram with three characters. This divides a sentence into sequences of three characters and calculates how many of those three-character words match.[39] The second index calculates how many changes have to be made on one name to make it identical to the other, normalized by the difference in the two names' length.

We define a "perfect" match between a lease in CompStak and an establishment in D&B if both observations belong to the same building and the similarity score between the two is the highest among all potential establishment matches in at least one of the indexes. "Good" matches are defined when the similarity scores in both indexes are the highest based on establishment name, but we cannot definitively confirm the records correspond to the same building (based on street address). "Good" matches represent 23% of our estimating sample in Table 5. We also estimated Table 5 restricting the sample to "perfect" matches; results were robust.

B.2 Data sources and access

B.2.1 Proprietary data

Our two primary datasets used in the paper are proprietary and we are not at liberty to post or share the raw data. These include the Dun and Bradstreet establishment level data and the CompStak establishment level data. A third dataset, from Safegraph, uses cellphone data to measure foot traffic to points of interest (POI) as one of the controls in our models. The Academic Partnership Program with Safegraph also does not allow us to share the raw data.

[39] We use the Stata program matchit to apply this algorithm and apply log weights to the three-character words based on frequency which minimizes false positive matches when encountering words like Inc or Cor.

Establishment level Dun & Bradstreet data were obtained from the Syracuse University library. Syracuse University has a site license with Merge Intellect that makes these data available to members of Syracuse University. Other universities and institutions with similar license would have comparable access. CompStak provides information on commercial leases. These data were obtained by purchasing an individual user license from CompStak Inc (https://compstak.com/) and can be similarly purchased by others.[40] Access to the Safegraph data can be requested at https://www.safegraph.com/academics. All other data used in the paper is publicly available.

A complete list of data sources for information used in the paper is below. This is followed by a list of the variables and their data source used in the paper, including those included in robustness checks that are discussed but not tabled out.

B.2.2 Data Sources

1995 Street Tree Census. Department of Parks and Recreation, NYC Open Data. Downloaded on December 21, 2019. https://data.cityofnewyork.us/Environment/1995-Street-Tree-Census/kyad-zm4j

2005 Street Tree Census. Department of Parks and Recreation, NYC Open Data. Downloaded on December 21, 2019. https://data.cityofnewyork.us/Environment/2005-Street-Tree-Census/29bw-z7pj

2015 Street Tree Census. Department of Parks and Recreation, NYC Open Data. Downloaded on December 16, 2019. https://data.cityofnewyork.us/Environment/2015-Street-Tree-Census-Tree-Data/uvpi-gqnh

311 Service Requests from 2010 to Present. Department of Information Technology and Telecommunications (DITT), NYC Open Data. Downloaded on March 8, 2020. https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9

Commercial Leases – CompStak. Downloaded on August 1, 2021. https://compstak.com/.

Establishments Directory – Dun & Bradstreet. Downloaded on October 13, 2018 and updated on February 27, 2019. https://www.mergentintellect.com/index.php/search/index.

In-Service Alarm Box Locations. Fire Department, NYC Open Data. Downloaded on December 16, 2019. https://data.cityofnewyork.us/Public-Safety/In-Service-Alarm-Box-Locations/v57i-gtxb

[40] The CompStak data are populated by leasing agents who provide CompStak information on leases they have executed in exchange for being able to draw other leases from the CompStak database that may be helping in guiding a new client.

National Registry of Historic Places. NY State Parks, Recreation and Historic Preservation. Downloaded on March 8, 2020. https://www.nps.gov/subjects/nationalregister/index.htm

New York City Community Air Survey (NYCCAS). Downloaded on March 9, 2020. https://www1.nyc.gov/site/doh/data/data-sets/air-quality-nyc-community-air-survey.page

New York Police Department (NYPD) Complaint Data History, NYC Open Data. Downloaded on September 13, 2020. https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i.

New York Police Department (NYPD) Stop, Question and Frisk Data. Downloaded on October 7, 2020. https://www1.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page.

NYC Condom Availability Program - HIV condom distribution locations. Department of Health and Mental Hygiene, NYC Open Data. Downloaded December 20, 2019. https://data.cityofnewyork.us/Health/NYC-Condom-Availability-Program-HIV-condom-distrib/4kpn-sezh

Open Space – Parks. Department of Parks and Recreation, NYC Open Data. Downloaded on March 6, 2020. https://data.cityofnewyork.us/Recreation/Open-Space-Parks-/g84h-jbjm

PLUTO and MapPLUTO version 18v2. Department of City Planning. Downloaded on January 1, 2020. https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page

Rodent inspections. Department of Health and Mental Hygiene (DOHMH), NYC Open Data. Downloaded on March 8, 2020. https://data.cityofnewyork.us/Health/Rodent-Inspection/p937-wjvj

SafeGraph Cellphone Data. Downloaded on March 2, 2020. https://www.safegraph.com/.

Subway Entrances. Metropolitan Transportation Authority (MTA), NYC Open Data. Downloaded on March 8, 2020. https://data.cityofnewyork.us/Transportation/Subway-Entrances/drex-xx56

Table B-1: Data Sources by Variable

Variable Groups	Variable	Definition	Source
Dependent variables	Neighborhood property crimes	Total number of property crimes occurred in the grid square. Property crime refers to burglary, grand larceny, petit larceny, theft of services or fraud.	New York Police Department (NYPD)
	Neighborhood police stops	Pedestrian stops made by the NYPD under the Stop-Question-Frisk policy. Only discretionary stops are included. Those pedestrian stops prompted by 911 calls or ongoing investigations are not considered.	New York Police Department (NYPD)
	Establishment input cost per dollar sold	Total expenditures on space and labor divided by the level of sales at the establishment level. See expression (2.9).	Dun & Bradstreet and CompStak Inc.
Baseline Controls	Distribution and level of employment	Total employment and employment in each selected industry (SIC)	Dun & Bradstreet
	Spatial Concentration Measures (*G*)	Spatial *G* for selected industries based on employment and sales	Dun & Bradstreet
	Monthly visits to POI	Average monthly visits to each POI in a grid square based on cellphone data	SafeGraph
	Number of Trees	Total number of trees on the street based on 2015 Street Tree Census	Department of Parks and Recreation
	Average Building Age	Average age of buildings across the grid square.	MapPLUTO
	Average Building Assessed Value	Average assessed value of the building in the grid square	MapPLUTO
	Overlapping Police Precincts	Grid square overlaps multiple police precincts	MapPLUTO
	Share of Residential Units Within Bldgs.	Share of all units in the grid square that are residential	MapPLUTO
	Neighborhood Sales per Worker	Total sales of single-site establishments in the grid square over total employment	MapPLUTO
Establishment Characteristics	Neighborhood Average Market Risk	Marketing Pre-screen Ranking: predicts the likelihood of a company to pay bills on-time. Ranges from 1 to 5, being 1 indicates most likely to pay	Dun & Bradstreet
	Neighborhood Avg. Establishment Age	Establishment Age = 2019 – Founding Year	Dun & Bradstreet
Zoning Restrictions	Share Special District	Share of lots in the grid located in special purpose districts.	MapPLUTO
	Share Commercial allowed in Residential	Share of lots in the grid that are allow for commercial overlay within a residential zoning district	MapPLUTO
	Share Multiple Zoning	Share of lots in the grid that are between multiple zoning features.	MapPLUTO
	Average Density Allowed by Residential Zoning	For lots in a residential district they are assigned a code from R1-1 to R10H, where the higher the number immediately after R the higher the density or intensity of land use permitted. We calculate the average of that number across the grid.	MapPLUTO
	Average Density Allowed by Commercial Zoning	For lots in a commercial district they are assigned a code from C1-6 to C8-4, where the higher the number immediately after C the higher the density or intensity of land use permitted. We calculate the average of that number across the grid.	MapPLUTO
	Share buildings with height restriction	Share of lots in the grid that are in a limited height district	MapPLUTO
	Average residential FAR	Maximum allowable residential floor area ratio across the grid	MapPLUTO
	Average commercial FAR	Maximum allowable commercial floor area ratio across the grid	MapPLUTO
Distance to Landmarks	Distance Central Park	Distance between the grid centroid and Central Park	Department of Parks and Recreation
	Distance Nearest Park	Distance from grid centroid to nearest park	Department of Parks and Recreation
	Distance Nearest Subway	Distance from grid centroid to nearest subway entrance (0 if entrance inside grid)	Metropolitan Transportation Authority
	# Subway Entrances	Number of subway entrances inside the grid square	Metropolitan Transportation Authority

Table B-1: Data Sources by Variable (continued)

Variable Groups	**Variable**	**Definition**	**Source**
Amenities	Average PM 2.5	Annual average fine particulate matter < 2.5 microns (2018), 300 mt resolution	Community Air Survey Air Pollution
	Ln(Reported rat problems)	Total 2018 rodent inspections that resulted in active rat signs.	Department of Health and Mental Hygiene
	Ln(Failed rodent inspections)	Total 2018 rodent inspections that did not pass the inspection.	Department of Health and Mental Hygiene
	Ln(Complaints about traffic lights)	311 Complaints (requests) related to traffic signal condition	Department of Information Technology and Telecommunications (DITT)
	Ln(Complaints about streetlights)	311 Complaints (requests) related to street light condition	DITT
	Newly planted trees 2005-2015	Difference between 2005 and 2015 Tree Census	Department of Parks and Recreation
	Historic Places and Landmarks	Historic places registered before 2018 to the National Register of Historic Places	NY State Parks, Recreation and Historic Preservation
	Public alarm boxes on the street	Fire alarm boxes in the grid: includes Emergency Reporting System (ERS) and Box Alarm Reporting System (BARS)	Fire Department
	Active Sites: HIV testing and condom distribution locations	Active venues distributing free safer sex products under the NYC Condom Availability Program – HIV.	Department of Health and Mental Hygiene
Building with Irregular shape and Tax exemptions	Share of buildings with irregular shape	Share of lots in the grid that have an irregular shape	MapPLUTO
	Share of buildings that are tax exempt	Share of lots in the grid that have at least 20% of their assessment value exempt of property tax.	MapPLUTO