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

Stuart S. Rosenthal  
Maxwell Advisory Board Professor of Economics  
Department of Economics and Center for Policy Research  
Syracuse University Syracuse, New York, 13244-1020  
Email: ssrosent@syr.edu

Joaquin A. Urrego  
Haas School of Business  
University of California – Berkeley, Berkeley, CA, 94720-1900  
Email: jaurrego@berkeley.edu

March 6, 2022

* 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 crime-deterrent 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. Point-specific 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 crime and spatial concentration shed light on mechanisms. Increasing neighborhood concentration of retail outlets by one standard deviation reduces property crime and police stops by 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 spatial patterns of retail outlets on property crime.\(^1\) Retail outlets have valuable inventory that attracts crime, but also draw crowds of shoppers. If concentrating retail establishments at the street level amplifies the effect of eyes on the street – a type of neighborhood-level external economies of scale – a simple model suggests that this should deter crime and reduce public and private investment in anti-crime measures, with benefits capitalized into higher retail rent.\(^2\) Moreover, this may occur not only because shoppers are crowded into smaller areas, but also by making it possible to observe multiple store fronts from a single location (referred to below as crowding and visibility effects, respectively). These and other predictions are confirmed using point-specific data for New York City.\(^3\)

The potential for crime deterrent effects of spatial concentration of retail establishments to yield substantial savings 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

---

\(^1\) 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).


NRF estimates suggest 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 a one standard deviation increase in block-level concentration of retail activity would reduce property crime by 8.5%. That represents a total savings among NYC retailers of roughly $117 million.

To frame our analysis, we develop a conceptual model built around the assumption that spatial concentration of retail establishments amplifies the effect of public and private protection against crime.\(^7\) Comparative statics then indicate that spatial concentration should reduce equilibrium levels of crime while lowering investment in public and private protection. In the empirical work that follows, we use multiple strategies to identify these relationships. This includes organizing geographically granular data into small neighborhood units, extensive controls that characterize neighborhood attributes, and various forms of differencing, details of which are described later. A threat to identification is that even after controlling for the many model features, in high crime neighborhoods retailers may seek to concentrate to gain better protection. Should that occur, we argue that our estimates will understate the crime deterrent effect of spatial concentration.

A second part of our model focuses on rent capitalization. Starting from the retailer’s profit function, we show that with competitive markets, equilibrium factor input costs per dollar of sales (based on space rented and labor) are equal to industry markup over wholesale cost, where markup allows for the cost of inventory lost to crime. Neighborhood spatial concentration reduces inventory lost to crime, savings from which are capitalized into higher local rent. A possible threat to identification in this part of the model is that spatial concentration also creates shopping externalities that likely increase sales for a

---

\(^5\) 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).


\(^7\) Our model is also implicitly based on the idea that criminals trade off potential return and costs in the spirit of Becker (1968). See Freedman and Owens (2016) for recent related empirical evidence.
given amount of space and labor, profits from which will also be capitalized into higher rent. However, because crime deterrence affects cost while shopping externalities affect sales, here too we argue that our estimates provide a lower bound on the rent capitalization effect of crime deterrence.8

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 (approximately a 4 to 7 minute walk), each of which is treated as a separate neighborhood. We then omit predominantly residential neighborhoods from our estimating samples (in a 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 using the location of each establishment (within roughly 3 feet). 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 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

---

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).

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 concerns about these issues. Details are provided later in the paper.

11 These findings are obtained from negative binomial regressions for counts of property crime and police stops. Negative binomial count models help to address overdispersion in our dependent variables that arise from zero values in many grid squares. Similar results were obtained using OLS.
includes the level and composition of employment in the neighborhood, controls for 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 help to shed light on crowding and visibility as underlying mechanisms. We compare crime rates at night to those during the day and also crime rates throughout 2018 to the first two weeks of the NYC COVID-19 lockdown (March 22nd – April 5th, 2020).12 Crowds are greatly diminished at night and during the lockdown for reasons unrelated to crime, and those shifts should weaken the effect of eyes on the street. In some models, we also split property crime into different sub-categories (petit larceny, grand larceny plus burglary, theft of services plus fraud) in addition to considering effects on robbery and auto theft. Evidence suggests that both crowding and enhanced visibility help to explain why spatial concentration of retail activity deters crime.

In a separate set of models, we measure spatial concentration in three different ways, based on spatial patterns of employment, the location of store fronts, and sales. We argue that the first measure is especially effective at capturing crowding effects, the second targets visibility, and the third is more of a placebo check having conditioned on the first two. Once again, results 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. 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

---

12 Later in 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](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.
protection measures that would discourage retail shoppers. Both effects should reduce the crime deterrent
effect of retail spatial concentration on wholesale establishment rent, which is what we find.

To establish these and related results, the next section presents our conceptual model. Section 3
describes the data and summary statistics. Section 4 presents the results, and Section 5 concludes.

2. Theoretical Framework

This section develops our conceptual model that motivates the empirical work to follow. The first
part considers the effect of spatial concentration of retail activity on equilibrium patterns of property
crime and police protection. The second part examines rent capitalization.

2.1 Equilibrium protection and the level of crime

We begin by assuming that all stores in a neighborhood \( j \) have identical valued inventory, \( I \), some
of which is stolen while the rest is sold,

\[
I_j = I_{\text{stolen}}(P_j) + I_{\text{sold}}(P_j).
\]  

(2.1)

In (2.1), stolen inventory declines with the quality of protection services enjoyed by each store in the
community, \( P_j \), where \( P_j \) is generated based on a Cobb-Douglas production technology,

\[
P_j = G_j P_u j^\alpha P_r j^\beta \quad \text{with } \alpha + \beta \leq 1.
\]  

(2.2)

In (2.2), \( P_u j \) and \( P_r j \) denote public and private expenditures on security, respectively, where \( P_u j \) includes
neighborhood police patrols, something we examine in the empirical work to follow, while \( P_r j \) includes
privately funded measures specific to each store, as with security alarms, door locks, surveillance
cameras, etc. Also included as an argument in (2.2) is the spatial concentration of retail activity in
neighborhood \( j \), denoted by \( G_j \). This enters as a Hicks neutral shift factor that amplifies the productivity of
public and private investment in security for reasons described in the Introduction and as motivated by
previous literature on “eyes on the street” (e.g. Chang and Jacobson, 2017; McMillen et al., 2019; Gonzalez and Komisarow, 2020).\(^{13}\)

To solve for the efficient mix of public and private investment in protection, suppose that there are \(N_j\) retail outlets in neighborhood \(j\), each of which pays an equal tax share to finance the cost of public protection. Establishments incur the following expenditures to support private and public security,

\[
\text{exp}_j = P r_j + \frac{1}{N_j} P u_j ,
\]

where \(1/N_j\) is the price of public protection, and the price of private protection is normalized to 1.

In the simplest setting, local government acts as a social planner and chooses \(P u_j\) and \(P r_j\) to minimize protection costs for each store while providing an efficient level of protection, \(P_j^*\).\(^{14}\) From (2.2) and (2.3), the Lagrangian is given by,

\[
\mathcal{L} = P r_j + \frac{1}{N_j} P u_j + \eta \left( P_j - G_j P u_j^\alpha P r_j^\beta \right) .
\]

Taking first-order conditions and rearranging, the efficient levels of public and private investment in protection for a given level of \(P_j\) are,

\[
P u_j^* = \left[ \frac{P_j}{G_j} \right]^{\frac{\beta}{\alpha + \beta}} N_j^{-\frac{\beta}{\alpha + \beta}} \left( \frac{\alpha}{\beta} \right)^{\frac{\alpha}{\alpha + \beta}}
\]

\[
P r_j^* = \left[ \frac{P_j}{G_j} \right]^{\frac{1}{\alpha + \beta}} N_j^{-\frac{\alpha}{\alpha + \beta}} \left( \frac{\alpha}{\beta} \right)^{\frac{\alpha}{\alpha + \beta}}
\]

Dividing, for each store the efficient ratio of public to private investment in protection is,\(^{15}\)

\[
\frac{P u_j}{P r_j} = N_j^{\frac{\alpha}{\beta}}
\]

\(^{13}\) 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 later in the paper. 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 \(P u_j\) for a given \(P r_j\), and similarly, that the private sector chooses an optimal \(P r_j\) for a given \(P u_j\). Expressions (2.5a) and (2.5b) then describe the corresponding reaction functions having substituted in \(P_j\) from (2.2). This yields the same combination of \(P u_j\) and \(P r_j\) as in (2.6) and also \(P_j^*\) as in (2.9) below.

\(^{15}\) In (2.6), the equilibrium ratio of public to private protection does not depend on \(G_j\) because \(G_j\) is specified as a Hicks neutral shifter above. \(N\) increases the equilibrium ratio of \(P u\) to \(P r\) because it lowers the price of public protection relative to private protection measures.
while the cost of providing $P_j$ protection to each store, $C_{P,j}$, is obtained by substituting (2.5a) and (2.5b) into (2.3). Grouping the production parameters into a positive constant $\kappa$, and solving,\(^{16}\)

$$C_{P,j} = \kappa \left[ \frac{\lambda_j}{G_j} \right]^{\frac{1}{\alpha + \beta}} \frac{1}{N_j}^{\frac{\alpha}{\alpha + \beta}}.$$  \hspace{1cm} (2.7)

To determine the efficient level of protection, we next characterize the deterrent effect of protection and the related amount of inventory that is stolen from each store. This is expressed in (2.8) as,

$$I_j^{sold} = I_j(1 - P_j^{-\lambda}) \text{ where } \lambda > 0 \text{ and } P_j > 1.$$ \hspace{1cm} (2.8)

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

Subtracting (2.7) from (2.8) yields a measure of the net gain to each store from its investment in private and public protection against property crime. Taking first order conditions and rearranging, each store in the community receives a level of protection given by,\(^{17}\)

$$P_j^* = \lambda^{\frac{\alpha + \beta}{\alpha + \beta} \frac{1}{1 + \lambda(\alpha + \beta)} G_j^{\frac{\alpha}{\alpha + \beta}} N_j^{\frac{\alpha}{\alpha + \beta}}}.$$ \hspace{1cm} (2.9)

Inventory lost to crime is then obtained by substituting (2.9) into (2.8) and using the identity in (2.1),

$$I_j^{stolen} = \left( \frac{1}{\lambda^{\frac{\lambda(\alpha + \beta)}{\alpha + \beta}}} \right) \frac{\lambda}{N_j^{\frac{\lambda}{\alpha + \beta}}} \left( \frac{1}{N_j} \right)^{\frac{\lambda}{\alpha + \beta}} \left( \frac{1}{G_j} \right)^{\frac{\lambda}{\alpha + \beta}} N_j^{\frac{\lambda}{\alpha + \beta}}.$$ \hspace{1cm} (2.10)

In (2.10), as $\lambda$ becomes increasingly large, criminal skill declines, deterrence increases, and inventory stolen goes to zero. In this instance, investment in protection converges to its lower bound of 1 in (2.9). Also evident from (2.10), higher $G_j$ reduces the amount of inventory stolen. This is because spatial concentration among retailers lowers the cost of protection in (2.7) while increasing equilibrium protection in (2.9). Expression (2.10) further highlights that crime increases with the level of retail

\[^{16}\kappa = \left( \frac{\alpha}{\beta} \right)^{\frac{\alpha}{\alpha + \beta}} + \left( \frac{\beta}{\alpha + \beta} \right)^{\frac{\beta}{\alpha + \beta}}.\]

\[^{17}\text{The first order condition requires that } \lambda_j P_j^{-(\lambda + 1)} - \kappa \frac{1}{\alpha + \beta} G_j^{\frac{1}{\alpha + \beta}} N_j^{\frac{1}{\alpha + \beta}} P_j^{\alpha + \beta - 1} = 0.\]
activity, as measured by inventory, and consistent with previous literature (Lee and Alshalan, 2005; Tillyer and Walter, 2019).18

2.2 Retail rent capitalization

We next consider the effect of inventory lost to crime on retail rent. Suppose initially that all retail establishments are identical. This assumption is relaxed later in the section and in the empirical work. We also defer discussion of shopping externalities to the end of this section where we show that shopping externalities cause our model to underestimate the effect of crime deterrence on rent.

2.2.1 Homogenous establishments

Each retail establishment sells \( q \) units of merchandise at \( p \). Retailers 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 \( c \). Retail product price, wage, and inventory cost (\( p \), \( w \), and \( c \), respectively) are determined at the metropolitan level and do not vary across neighborhoods. The share of inventory lost to crime shrinks with neighborhood spatial concentration of retail activity, \( G_j \). Output \( q \) is produced using labor and space, the productivity of which are amplified by shopping externalities that also increase with \( G_j \).

Collecting terms, profit for an establishment in neighborhood \( j \) is given by,

\[
\pi = pq(L_j, S_j; G_j) - wL_j - r(G_j)S_j - C(c, G_j)q(L_j, S_j; G_j),
\]

In (2.11), observe that \( C(c, G_j) \) is the cost of inventory for each unit sold where \( C \) increases with \( c \) and is inversely related to \( G_j \). Because \( G_j \) enhances both crime deterrence and shopping externalities, higher \( G_j \) should be capitalized into higher local rent, \( r_j \).

We define the cost of inventory \( C(c, G_j) \) as,

---

18 Notice also that as \( \lambda \) shrinks to zero in (2.10), \( I_j^{stolen} \) 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.
\[ C(c,G_j) = c + cP^{-\lambda}(G_j) , \]  

where \( c \) is the wholesale price per unit of inventory as above, \( P \) is the level of protection as defined in (2.9), and \( P^{-\lambda}(G_j) \) is the share of inventory lost to crime as defined in (2.8). Substituting (2.12) into (2.11) and setting \( \pi = 0 \) with competitive markets, expenditure on space per dollar sold can be written as,

\[ \frac{r(g)k_j}{pq(g)} = \theta - \frac{wL_j}{pq(g)} - \gamma(G_j) , \]

where \( \theta = \frac{p-c}{p} \) and \( \gamma(G_j) = \frac{cP^{-\lambda}(G_j)}{p} \).

In (2.13), \( \theta \) is the percentage markup of retail to wholesale price and is common across establishments in the metropolitan area. The term \( \gamma(G_j) = \frac{cP^{-\lambda}(G_j)}{p} \) is a neighborhood-specific markup that allows for the share of inventory lost to crime. The remaining term, \( \frac{wL_j}{pq(g)} \), is labor cost per dollar sold. In our data we observe \( \frac{L_j}{pq(g)} \) 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 \( \frac{wL_j}{pq(g)} \) 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 (2.13) as,

\[ \frac{r(g)k_j+wL_j}{pq(g)} = \theta - \gamma(G_j) , \]

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

### 2.2.2 Heterogeneous establishments

In this section we highlight three sources of heterogeneity that affect the dependent variable in (2.14). One is that companies belong to different industries, \( k = 1, \ldots K \), each of which may have its own markup, \( \theta_k \). 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 could 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 $\theta_i$, for $i = 1, \ldots I$ establishments. Collecting terms and suppressing the $i$ subscripts on $S$ and $L$ to simplify, (2.14) becomes,

\[
\frac{r(G_j)S_j + wL_j}{pq(G_j)} = \theta_k - \gamma(G_j) + b z_j + \theta_i . \tag{2.15a}
\]

Our primary goal with (2.15a) is to measure the effect of $G_j$ on $\gamma(G_j)$. We do this in three ways. In the first approach, we estimate (2.15a) using neighborhood fixed effects to measure $\gamma(G_j)$. These are then regressed on $G_j$ to summarize the average relationship between $\gamma(G_j)$ and $G_j$. A second, more general approach is to estimate (2.15a) using Robinson’s (1988) partial linear model in which $\gamma(G_j)$ is estimated nonparametrically while specifying a parametric structure for the other model terms. Results from both approaches indicate that $\gamma(G_j)$ is approximately linear in $G$. Partly for that reason and to simplify presentation, in a third approach we impose a linear approximation on $\gamma(G_j)$. Taking a first order Taylor expansion of $\gamma(G_j)$ around $\bar{G}$, the sample mean of $G$, (2.15a) becomes,

\[
\frac{r(G_j)S_j + wL_j}{pq(G_j)} = \tilde{\theta}_k - \gamma'(\bar{G})G_j + b z_j + \theta_i , \tag{2.15b}
\]

where $\tilde{\theta}_k = \theta_k - \gamma(\bar{G}) + \gamma'(\bar{G})\bar{G}$. Notice that if $\gamma(\bar{G})$ is linear then $\gamma(\bar{G}) = \gamma'(\bar{G})\bar{G}$ and $\tilde{\theta}_k = \theta_k$ so that $\tilde{\theta}_k$ equals industry markup. Also, and of primary interest, with further manipulation, the coefficient on $G_j$ in (2.15b) can be written as $\gamma'(\bar{G}) = -a P^{-2} \bar{G}^{-1}$, where $a$ is a positive constant.\(^{19}\) This measures the marginal effect of $G$ on the cost of inventory lost to crime per dollar sold evaluated at $\bar{G}$.

---

\(^{19}\) Recall from (2.13) that $\gamma(G_j) = \frac{cP^{-\lambda}(G_j)}{p}$. Differentiating with respect to $G$,

\[
\gamma'(G_j) = -\frac{c}{p} \lambda P_{-\lambda}(1+\lambda) \frac{\partial P_j}{\partial G_j} = -a P_j^{-3} G^{-1} < 0 , \tag{N.1}
\]

where $a = \left(\frac{c}{p}\right) \frac{\lambda}{1+\lambda(a+\beta)} > 0$ and from (2.9) $\frac{\partial P_j}{\partial G_j} = \frac{1}{1+\lambda(a+\beta)} P_j G^{-1}$. Taking a first order Taylor expansion of $\gamma(G_j)$ around $\tilde{G}$ and substituting (N.1) into (2.15a) gives $\gamma'(\tilde{G}) = -a P_j^{-3} \tilde{G}^{-1}$. 

10
In (2.15a) and (2.15b), note 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 $\gamma(G_j)$ and $hz_j$ should be capitalized into higher local rent. Entrepreneur skill, as captured by $\theta_i$, should instead increase profit. In the empirical work to follow, we proxy for $\theta_i$ 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.

2.2.3 Lower bound on rent capitalization from crime deterrence

As suggested earlier, the model above likely 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 also 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 likely 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.\(^{20}\)

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 (2.15b) 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:

$$\frac{r(\alpha_j)S_j}{r(\alpha_j)S_j + \omega L_j} < \frac{q(\alpha_j)}{q(\alpha_j)}$$

\(^{(2.16)}\)

\(^{20}\) 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 (2.15b). Instead, it is the productivity advantages from shopping externalities that enhance net profit. See Koster et al. (2019) for related discussion.
where $r'(G_j)$ and $q'(G_j)$ are derivatives with respect to $G_j$.

Note now that $wL_j > 0$ so that \[ \frac{r(G_j)s_j}{r(G_j)s_j + wL_j} < \frac{r(G_j)}{r(G_j)}. \] Also, $\frac{r(G_j)}{r(G_j)}$ and $\frac{q(G_j)}{q(G_j)}$ are approximately equal to $\%\Delta r$ and $\%\Delta q$, respectively. A sufficient condition for (2.16) to hold is that productivity gains from shopping externalities have a smaller percentage effect on rent $r$ than on sales $q$. Moreover generally, (2.16) will hold provided that $\%\Delta r$ is not substantially larger than $\%\Delta q$ since \[ \frac{r(G_j)s_j}{r(G_j)s_j + wL_j} < \%\Delta r. \] This condition is undoubtedly met. Shopping externalities, therefore, shrink the dependent variable in our capitalization expressions while crime deterrence has the opposite effect. For this reason, our model will tend to understate the rent capitalization effect of crime deterrence.21

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 A. 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 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. Grid squares are independent of administrative boundaries and each is treated as a separate neighborhood. It is worth

\[21\text{ Note also that if capitalization of } G \text{ into higher rent prompts retailers to substitute } L \text{ for } S, \text{ related cost savings will cause our model to further understate capitalization effects from crime deterrence.}\]
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 on how this was determined are discussed after other features of the data have been described.

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,

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

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 \( \omega_{ie} \), the weight placed on employment as distance, \( d_{ie} \), increases from establishment \( i \). We adopt the following weight function:

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

This function sets \( \omega_{ie} \) to 1 for all establishments within 250 feet of \( i \). For distances between 250 to 1,000 feet from \( i \), \( \omega_{ie} \) is assumed to decline with distance at rate \( 1/(d_{ie} - 250)^{0.7} \), where the exponent 0.7 was chosen to set \( \omega_{ie} \) to roughly 1% at 1,000 feet. Beyond 1,000 feet, \( \omega_{ie} \) 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 \( G \).\(^{22}\) Note further that specified as above, \( G_i \) is measured separately for each establishment and varies within a given neighborhood.

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,

\[
\tilde{G}_j = \frac{1}{sd(\bar{G}_j)} \left( \bar{G}_j - \frac{1}{m} \sum_{j=1}^{m} \bar{G}_j \right).
\]

(3.3a)

In this expression, \( \bar{G}_j = \frac{1}{nj} \sum_{i=1}^{nj} G_i \) 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,

\[
\tilde{G}_i = \frac{1}{sd(\bar{G}_i)} \left( \bar{G}_i - \frac{1}{I} \sum_{i=1}^{I} \bar{G}_i \right).
\]

(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 store fronts at once). As described in the Introduction, in some models we also add a measure of \( \tilde{G} \) based on the spatial concentration of store fronts, \( \tilde{G}_{Stores} \), 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:

\(^{22}\) 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 \( \omega_{te} \).
\[
G_{\text{Ret,Stores}} = \sum_e \omega_{ie}(d_{ie}) / n(d_{ie}),
\]
(3.4)

where \( n_i \) is the number of establishments within a given distance of store \( d_{ie} \), \( \omega_{ie} \) is defined as in (3.2), and \( G_{\text{Stores}} \) is normalized as in (3.3a). As a placebo check, we also consider a third measure of \( G \) based on sales, denoted as \( G_{\text{Sales}} \) (calculated in the manner as \( G_{\text{Emp}} \) except using sales). Controlling for \( G_{\text{Emp}} \) and \( G_{\text{Stores}} \), we do not expect \( G_{\text{Sales}} \) to affect eyes on the street.

When measuring the level and composition of neighborhood employment, we use all establishments in the neighborhood. When measuring \( G \), however, we use only single site firms which account for 95 percent of all establishments in NYC. We do this to target external economies of scale that arise from clustering companies together and which have potential to enhance the eyes on the street phenomenon. To clarify why, consider the following.

Single-site establishments tend to be small and have limited resources to invest in private protection, both because they are small and because they do not belong to large multi-site firms that might otherwise help to support loss protection measures at individual outlets.\(^{23}\) In an NBC news report (December 2, 2021), 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.”


Imagine now two neighborhoods, one with a single big-box retail outlet and the other with many small independent retail companies clustered together. Employment would be concentrated in both instances but for different reasons that affect interpretation of \( G_{\text{Emp}} \). In the first neighborhood, high \( G_{\text{Emp}} \) proxies for high \( P_{ij} \) in (2.2) since big-box stores tend to invest heavily in private protection. In the second neighborhood, high \( G_{\text{Emp}} \) proxies for external economies of scale that are expected to enhance eyes on the street. Including large companies when measuring \( G_{\text{Emp}} \) confounds these two cases. Restricting the

\(^{23}\) Among all industries combined, the number of workers in NYC at the 75\(^{th}\), 95\(^{th}\) and 99\(^{th}\) size percentiles are: 3, 10, and 40 for single-site companies, and 26, 150 and 500 for establishments belonging to multi-site firms.
sample used to measure $\tilde{G}_{Emp}$ to just the single-site companies, however, ensures that $\tilde{G}_{Emp}$ can only be large when many establishments are grouped together.$^{24}$

### 3.2 Data

A complete list of measures used in our regression models and their sources is provided in Appendix A. 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 of the incident. 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, 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.

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.$^{25}$ It is also worth noting that SQF has been

---

$^{24}$ Using only single-site companies also allows us to use the same underlying sample when measuring $\tilde{G}_{Emp}$, $\tilde{G}_{Stores}$ and $\tilde{G}_{Sales}$ since establishment sales are only reliably reported for single-site companies.

$^{25}$ 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.
controversial because of concerns about racial bias. Although the SQF policy was initiated in NYC prior to 2000, it became widely used during the Bloomberg administration (2002-2013) when stops grew from 97,296 in 2002 to over 500,000 in 2012 (Evans et al., 2014). By 2012, numerous newspaper articles had reported on allegations of excessive force being used when stopping African Americans and many court cases had followed. In response, by 2016 SQF stops were substantially reduced and better targeted at potential criminal activity as evidenced by sharply higher arrest rates upon making a stop (Urrego, 2021). In 2016, 2017 and 2018 the number of SQF stops were just 12,053, 11,204 and 11,008, respectively.\(^{26}\)

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 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 and are current as of that time.

### 3.2.2 Commercial rent

Commercial lease data were obtained from CompStak and are used to estimate the rent capitalization models. For each lease, CompStak reports effective rent per square foot of space leased,

---

\(^{26}\) 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).
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 (2.14), (2.15a) and (2.15b). 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 attributes

Numerous measures were obtained from various New York City government agencies and coded up to the neighborhood level. This includes data obtained from the New York City Department of City Planning and the Department of Finance MapPLUTO 18v2 map. This map provides detailed information on the attributes of each tax lot in NYC. This includes information about the building situated on the lot, zoning, tax assessments, and many 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.

We also control for foot traffic to Points of Interest (POI) using cellphone data obtained from SafeGraph. SafeGraph defines over 110,000 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.27

Our more robust models include 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). In a robustness check, an additional 25 neighborhood attributes are added to the crime and police stop models (in Appendix B).

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 $\tilde{G}_{\text{Emp}}$ 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 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.

27 We also calculated spatial $G$ measures for POI. Correlation between that measure and spatial concentration of retail employment was just 8%. Including spatial concentration of POI visits also had no effect on the coefficients on the other model estimates and was dropped from the regressions to simplify specification and discussion.
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 variance of the outcome measures exceed their means.28

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.

---

28 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.
Also in Table 1b, the distribution of the ratio of retailer cost of space per dollar of sales often exceeds one and especially so for non-inventory costs per dollar sold (which includes labor costs). This is to be expected given the heavy reliance of small businesses on financing. It also likely reflects the high tendency of businesses to fail in their first few years, presumably because they are not able to generate sufficient revenue to cover costs. 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. We anticipate therefore that the cost/sales ratio will often exceed one as in Table 1b.

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.\(^{29}\) Recall that $\tilde{g}_{Emp}$ is normalized to have mean zero and standard deviation of 1 so that a one-unit change in $\tilde{g}_{Emp}$ equals one standard deviation across neighborhoods. Columns 1-4 present estimates of property crime while columns 5-8 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 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 5 (for police stops), the sample includes all 5,461 neighborhoods in which there is at least some

\(^{29}\) Marginal effects from a negative binomial regression can be calculated using the expression, $\exp(\beta_x \Delta x)$, where $\Delta x$ represents the change in a control variable and $\beta_x$ is its corresponding coefficient. This expression gives the change in counts. $\beta_x$ 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 $\beta_x$.\(^{29}\)
presence of both retail and non-retail employment. In columns 2 and 6, 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 -0.087.30

Columns 3 (for crime) and 7 (for police stops) expand the set of controls to include employment shares and spatial concentration for retail, finance insurance and real estate (FIRE), manufacturing and the service sector. Columns 4 and 8 add in controls for additional neighborhood features. These include 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. Columns 4 and 8 also include average monthly visits to POI in the grid square to further control for the level of neighborhood activity. Comparing across columns, notice that there is little change in the coefficients on retail employment and spatial concentration as one moves from left to right. Once again, the estimates of interest are robust.

The magnitude of the coefficients is also of interest. In columns 4 and 8, doubling grid square aggregate employment increases property crime and police stops by 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 that property crime and police stops increase with the overall level of foot traffic in a neighborhood (as captured by POI visits) and especially so with an increase in business activity (as reflected in total employment).

30 Analogous results are obtained when controls are added for the service sector which is present in all neighborhoods where retail is present.
Also noteworthy in Table 2, 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 4, a one unit increase in $\bar{g}_{\text{Ret,Emp}}$ – 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 $\bar{g}_{\text{Ret,Emp}}$ 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. We recognize that a critique of this interpretation could be 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 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.

A final point worth noting is that business establishments may shy away from high crime areas and/or may spatially concentrate in such locations to gain better protection. To the extent that occurs, the estimates in Table 2 will tend to understate the effect of employment levels and spatial concentration on crime and police stops.
4.1.2 Robustness checks

Three different types of robustness checks are presented in Appendix B, in each case focusing on the models in columns 4 (crime) and 8 (police stops) of Table 2.

The first check adds 25 additional neighborhood level controls to the core models. These include 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 are robust to the additional controls.\(^3\) In a second robustness check, we measure \(G_{Emp}\) using both single-site establishments and those that belong to multi-site firms. In a third robustness check, we define neighborhoods as 3 by 3 configurations of grid squares. \(G_{Emp}\) is then measured as the sum of squared employment shares across the 9 squares (using single-site companies).

For all three alternate model designs, although the magnitude of some of the coefficients differs, the central results are always robust. This includes that (i) retail employment share has a disproportionate 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.

4.1.3 Mechanisms: crowding and visibility

Tables 3 and 4 use alternate strategies to shed light on crowding and visibility as mechanisms that may account for why spatial concentration of retail activity deters crime. In Table 3 we compare estimates

\(^3\) We also explored adding controls for the portion of a grid square that belongs to a Business Improvement District (BID) as BIDs may pool resources and invest in local security measures (see Faggio, 2021, for related work on the relationship between BIDs and crime). To do so, we experimented with different combinations of dummy variables for the share of a grid square that is within the boundaries of a BID, including 0, more than 0 and less than 25%, 25% to 50%, etc. In some specifications there was evidence of a positive relationship between BID presence and neighborhood crime. That could arise if BIDs form in high crime areas in part to help protect against crime (see Faggio, 2021, for similar results), in which case BID presence would be endogenous. We also found that BID presence did not have any discernible effect on police stops, and most important, including BID measures had no effect on the core estimates including those for spatial concentration. For these reasons, we chose not to include controls for BID presence in the crime and police stop models.
for different sample periods and types of crime that are likely to exhibit different sensitivity to the two mechanisms. In Table 4 we compare estimates for the three different measures of spatial concentration described earlier, $\tilde{C}_{Ret,Emp}$, $\tilde{C}_{Ret,Stores}$ and $\tilde{C}_{Ret,Sales}$. In both tables, we re-estimate columns 4 (for crime) and 8 (for police stops) from Table 2 having made the specified adjustment. 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.\footnote{\textbf{\textsuperscript{32}} During these two weeks, stores were largely closed for in-person visits but the NYPD was still largely fully staffed. As noted in the Introduction, not long after, the high rate of COVID-19 cases left the NYPD short staffed which may have reduced patrol activity.} 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.\footnote{\textbf{\textsuperscript{33}} 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.}

When considering the patterns in Table 3, crowding of course is reduced at night and during the lockdown.\footnote{\textbf{\textsuperscript{34}} SafeGraph cell phone data indicate that foot traffic in NYC fell by 60\% in April, 2020.} 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 shop lifting 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 offence. 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 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 4. 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 4 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 $\tilde{G}_{\text{Emp}}$, $\tilde{G}_{\text{Stores}}$, and $\tilde{G}_{\text{Sales}}$ for each of the four highlighted industries (retail, finance, manufacturing and service). Conditioning on all three measures at once, $\tilde{G}_{\text{Emp}}$ primarily targets crowding, $\tilde{G}_{\text{Stores}}$ proxies the potential to observe multiple storefronts from a single location (a feature of visibility), and $\tilde{G}_{\text{Sales}}$ 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 $\tilde{G}_{\text{Ret.Emp}}$ and $\tilde{G}_{\text{Ret.Sales}}$ is high, 54%. Correlation between $\tilde{G}_{\text{Ret.Emp}}$ and $\tilde{G}_{\text{Ret.Stores}}$, however, is just 14%, and correlation between $\tilde{G}_{\text{Ret.Sales}}$ and $\tilde{G}_{\text{Ret.Stores}}$ is 15%. These summary measures confirm that $\tilde{G}_{\text{Ret.Emp}}$, $\tilde{G}_{\text{Ret.Stores}}$ and $\tilde{G}_{\text{Ret.Sales}}$ contain different information.

Consider now Panel B of Table 4 which presents estimates of the crime and police stop models. Models are reported for the same mix of time periods as in Table 3. The exception is for the first two weeks of the COVID-19 lockdown for which there were not enough police stops to estimate the model. 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, $\tilde{G}_{\text{Ret.Emp}}$, $\tilde{G}_{\text{Ret.Stores}}$ and $\tilde{G}_{\text{Ret.Sales}}$. Coefficients on the other model controls are suppressed to conserve space.

Focusing on police stops first (columns 5-8), notice that $\tilde{G}_{\text{Ret.Emp}}$ reduces stops both during the day and at night, but $\tilde{G}_{\text{Ret.Stores}}$ has no effect in either period. This indicates that police stops are less prevalent in crowded areas but are apparently not so sensitive to visibility of storefronts.

The patterns for property crime (columns 1-4) reinforce those for police stops and also display differences. In this case, crowding as proxied by $\tilde{G}_{\text{Ret.Emp}}$, clearly deters property crime during the day and in the lockdown but as in Table 3, the effect of $\tilde{G}_{\text{Ret.Emp}}$ is reduced at night when crowds are mostly
absent. Along with estimates from the police stop models, this pattern appears to confirm that crowding enhances the crime deterrent effects of eyes on the street.

Different from the police stop models, notice also that visibility, as proxied by $\hat{G}_{\text{Store}}$, 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 suggests that visibility and the potential for police, other individuals, and/or security cameras to observe multiple storefronts at once also enhances eyes on the street and helps to deter property crime.

A last point to note in Table 4 is that for both property crime and police stops, the coefficients on $\hat{G}_{\text{Sales}}$ 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.\(^3\)\(^5\)

We begin by estimating (2.14) and (2.15a) 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 (2.14)) while Panel B controls for additional neighborhood and establishment attributes that may affect the dependent variable (as in (2.15a)). 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 $\gamma(G_{\text{Emp}})$ using Robinson’s (1988) two-step partial linear model drawing on the semipar routine in Stata (Verardi and Debarsy, 2012). This estimates the $\gamma(G_{\text{Emp}})$ function non-parametrically with optimal smoothing. As before, Panel A does not allow

---

\(^3\) For these models, we once again measure spatial concentration based only on employment, in part because of the more limited sample for the matched sample.
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 $\gamma(G_{\text{Retail},\text{Emp}})$ is approximately linear in $\tilde{G}_{\text{Retail},\text{Emp}}$. This last observation supports our remaining empirical exercise.

Table 5 reports estimates of expression (2.15b) in which we impose a linear approximation on the relationship between $\gamma(G_{\text{Retail},\text{Emp}})$ and $\tilde{G}_{\text{Retail},\text{Emp}}$. 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 (2.13)). 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. 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 (2.15b). 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 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.

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 from a combination of theft and expenditures to protect against crime, a substantial amount relative to profit margins that

36 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.

37 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.
average roughly 3%.\(^{38}\) Local public authorities also devote considerable resources to patrolling retail
districts. We show that concentrating retail establishments at the street level has potential to notably
reduce these costs by fostering external economies of scale that enhance the deterrent effect of eyes on the
street.

Using data for New York City, findings indicate that for a shift from the 25\(^{th}\) to the 75\(^{th}\) 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 at least 9.6\%. These estimates are robust to an extensive set
of controls that describe the attributes of small (0.2 by 0.2 square mile) neighborhoods across NYC. Our
estimates are also large enough to be important for local policy makers and business establishments. This
is especially so in that the primary threats to identification cause our models to underestimate the effect of
spatial concentration on crime and rent capitalization.

Additional analysis sheds light on mechanisms. We compare crime and police stops in daytime
versus nighttime and pre-pandemic versus COVID-19 lockdown. Crowds are greatly diminished at night
and during lockdown. Visibility is also reduced at night but not during the lockdown. Evidence based on
these and other features of the models suggest that crowding and visibility both contribute to eyes on the
street. In a complementary exercise, we create separate measures of spatial concentration based on
employment – which targets crowding – and proximity of storefronts – which targets the ability to
observe multiple stores from a single location. Once again, evidence suggests that crowding and visibility
both enhance eyes on the street.

Together the various models and estimates in our paper suggest that 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.

\(^{38}\) See discussion by the Small Business Resource Center at https://sbrc.employers.com/retail/whats-a-good-profit-
References


Figure 1a: Retail Employment in Included Neighborhoods

Figure 1b: Spatial Concentration of Retail Employment ($\hat{G}$) in Included Neighborhoods
Figure 1c: Retail Employment and Spatial Concentration South of Central Park

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, $\tilde{G}_{\text{RetailEmp}}$, 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 $\tilde{G}^a$

Panel A plots the estimated grid square fixed effects from expression (2.14). Panel B does the same but includes controls for other neighborhood and establishment attributes as in (2.15a) 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.

$^a$ Panel A plots the estimated grid square fixed effects from expression (2.14). Panel B does the same but includes controls for other neighborhood and establishment attributes as in (2.15a) 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 $-\gamma(\tilde{G})$.

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 $\gamma$, 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.

\[^a\text{Panel A plots the estimated } -\gamma(\tilde{G}) \text{ function from a partial linear model where non-inventory cost per dollar of sales depends, in addition to } \gamma, \text{ on subindustry fixed effects. Panel B adds other controls linearly to the specification in Panel A. Those are spatial concentration of other industries, visits to POI, share of residential units, number of trees and fixed effect for different age categories of retail establishments.}\]
Table 1a: Attributes of Included and Omitted Neighborhood Grid Squares

Panel A: Grids included in our Estimating Sample

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

Panel B: Grid Squares not included in our Estimating Sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs.</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>p10</th>
<th>p50</th>
<th>p90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of Establishments</td>
<td>2,727</td>
<td>40.34</td>
<td>43.42</td>
<td>5</td>
<td>33</td>
<td>82</td>
</tr>
<tr>
<td>Total Employment</td>
<td>2,727</td>
<td>214</td>
<td>526</td>
<td>19</td>
<td>107</td>
<td>444</td>
</tr>
<tr>
<td>Employment: Retail</td>
<td>2,727</td>
<td>31</td>
<td>99</td>
<td>0</td>
<td>9</td>
<td>68</td>
</tr>
<tr>
<td>Employment: Finance</td>
<td>2,727</td>
<td>11</td>
<td>64</td>
<td>0</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>Employment: Manufacturing</td>
<td>2,727</td>
<td>4</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Employment: Services</td>
<td>2,727</td>
<td>117</td>
<td>379</td>
<td>5</td>
<td>46</td>
<td>231</td>
</tr>
<tr>
<td>Share of Residential Units Within Buildings</td>
<td>2,708</td>
<td>0.86</td>
<td>0.30</td>
<td>0.210</td>
<td>0.988</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Table 1b: Summary Statistics

<table>
<thead>
<tr>
<th>Panel A: At the grid square level&lt;br&gt;</th>
<th>Obs.</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>p25</th>
<th>p50</th>
<th>p75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Property crime</td>
<td>3,506</td>
<td>29.06</td>
<td>54.78</td>
<td>6</td>
<td>14</td>
<td>31</td>
</tr>
<tr>
<td>Grand Larceny &amp; Burglary</td>
<td>3,506</td>
<td>9.76</td>
<td>16.37</td>
<td>2</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Petit Larceny</td>
<td>3,506</td>
<td>18.40</td>
<td>40.45</td>
<td>3</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>Theft of Services &amp; Fraud</td>
<td>3,506</td>
<td>0.90</td>
<td>2.13</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Robbery</td>
<td>3,506</td>
<td>2.57</td>
<td>3.48</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Auto Theft</td>
<td>3,506</td>
<td>0.95</td>
<td>1.18</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Police Stops</td>
<td>3,506</td>
<td>2.08</td>
<td>4.37</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>At Least 1 Police Stop (Stops &gt; 0)</td>
<td>3,506</td>
<td>0.49</td>
<td>0.50</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

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<sup>d</sup>

\[ G_{Retail.Emp} \]
\[ G_{Finance.Emp} \]
\[ G_{Manuf.Emp} \]
\[ G_{Service.Emp} \]

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

<table>
<thead>
<tr>
<th>Panel B: At the establishment level&lt;br&gt;</th>
<th>Obs.</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>p25</th>
<th>p50</th>
<th>p75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Concentration of Retail&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1,596</td>
<td>0.01</td>
<td>0.98</td>
<td>-0.57</td>
<td>0.02</td>
<td>0.65</td>
</tr>
<tr>
<td>Retailer Rent per sq foot (psf) per month</td>
<td>1,596</td>
<td>178.02</td>
<td>211.27</td>
<td>61.38</td>
<td>110.98</td>
<td>219.75</td>
</tr>
<tr>
<td>Retailer Employment</td>
<td>1,596</td>
<td>7.21</td>
<td>11.49</td>
<td>2</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Retailer Sales</td>
<td>1,596</td>
<td>471.918</td>
<td>1,438,367</td>
<td>89,625</td>
<td>150,000</td>
<td>289,590</td>
</tr>
<tr>
<td>Retailer Cost of space / Sales</td>
<td>1,596</td>
<td>2.95</td>
<td>4.03</td>
<td>0.55</td>
<td>1.38</td>
<td>3.59</td>
</tr>
<tr>
<td>Retailer Non-inventory cost / Sales&lt;sup&gt;e&lt;/sup&gt;</td>
<td>1,596</td>
<td>4.15</td>
<td>4.29</td>
<td>1.45</td>
<td>2.73</td>
<td>5.27</td>
</tr>
<tr>
<td>Wholesaler Cost of space / Sales</td>
<td>538</td>
<td>0.66</td>
<td>1.06</td>
<td>0.06</td>
<td>0.21</td>
<td>0.74</td>
</tr>
<tr>
<td>Wholesaler Non-inventory cost / Sales&lt;sup&gt;e&lt;/sup&gt;</td>
<td>538</td>
<td>1.44</td>
<td>1.41</td>
<td>0.43</td>
<td>0.97</td>
<td>1.89</td>
</tr>
</tbody>
</table>

<sup>a</sup> 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.

<sup>b</sup> The unit of analysis is a grid cell of 0.2 square miles across the New York City area.

<sup>c</sup> Panel B includes only establishments matched in the CompStak and Dun and Bradstreet files.

<sup>d</sup> Spatial concentration is calculated as described for \( G \) in the text.

<sup>e</sup> 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$^a$

<table>
<thead>
<tr>
<th>Employment</th>
<th>Property crime$^b$</th>
<th></th>
<th></th>
<th>Police Stops$^c$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log total employment</td>
<td>0.722***</td>
<td></td>
<td></td>
<td>0.554***</td>
<td>0.539***</td>
<td>0.394***</td>
<td>0.463***</td>
<td>0.349***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td></td>
<td></td>
<td>(0.022)</td>
<td>(0.031)</td>
<td>(0.037)</td>
<td>(0.041)</td>
<td>(0.056)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share of Employment: Retail</td>
<td>2.557***</td>
<td></td>
<td></td>
<td>2.307***</td>
<td>2.002***</td>
<td>2.411***</td>
<td>2.640***</td>
<td>2.017***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.202)</td>
<td></td>
<td></td>
<td>(0.165)</td>
<td>(0.254)</td>
<td>(0.279)</td>
<td>(0.373)</td>
<td>(0.359)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share of Employment: Finance</td>
<td>-</td>
<td></td>
<td></td>
<td>-0.451**</td>
<td>-</td>
<td>-</td>
<td>-0.574</td>
<td>-0.752</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td>(0.217)</td>
<td>-</td>
<td>-</td>
<td>(0.626)</td>
<td>(0.608)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share of Employment: Manufacture</td>
<td>-</td>
<td></td>
<td></td>
<td>-0.262</td>
<td>-</td>
<td>0.293</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td>(0.247)</td>
<td>-</td>
<td></td>
<td>(0.757)</td>
<td>(0.536)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share of Employment: Services</td>
<td>-</td>
<td></td>
<td></td>
<td>0.739***</td>
<td>-</td>
<td>0.941***</td>
<td></td>
<td>0.055</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td>(0.129)</td>
<td>-</td>
<td>(0.293)</td>
<td></td>
<td>(0.318)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log monthly visits to POI</td>
<td>-</td>
<td></td>
<td></td>
<td>0.516***</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.894***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td>(0.030)</td>
<td>-</td>
<td>-</td>
<td></td>
<td>(0.069)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial concentration$^d$</td>
<td>-0.097***</td>
<td></td>
<td></td>
<td>-0.085***</td>
<td>-0.150***</td>
<td>-0.087**</td>
<td>-0.062</td>
<td>-0.110***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td></td>
<td></td>
<td>(0.015)</td>
<td>(0.036)</td>
<td>(0.040)</td>
<td>(0.040)</td>
<td>(0.036)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retail ($\tilde{G}_{\text{Ret,Emp}}$)</td>
<td>-</td>
<td></td>
<td></td>
<td>0.005</td>
<td>-</td>
<td>-</td>
<td>0.081**</td>
<td>0.042</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td>(0.016)</td>
<td>-</td>
<td>-</td>
<td>(0.041)</td>
<td>(0.040)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finance ($\tilde{G}_{\text{Fin,Emp}}$)</td>
<td>-</td>
<td></td>
<td></td>
<td>0.007</td>
<td>-</td>
<td>-</td>
<td>-0.037</td>
<td>-0.020</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td>(0.018)</td>
<td>-</td>
<td>-</td>
<td>(0.045)</td>
<td>(0.041)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manufacture ($\tilde{G}_{\text{Manf,Emp}}$)</td>
<td>-</td>
<td></td>
<td></td>
<td>0.018</td>
<td>-</td>
<td>-</td>
<td>(0.120)</td>
<td>(0.020)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td>(0.018)</td>
<td>-</td>
<td>-</td>
<td>(0.042)</td>
<td>(0.046)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Services ($\tilde{G}_{\text{Serv,Emp}}$)</td>
<td>-</td>
<td></td>
<td></td>
<td>-0.029</td>
<td>-</td>
<td>-</td>
<td>-0.239**</td>
<td>-0.137***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td>(0.019)</td>
<td>-</td>
<td>-</td>
<td>(0.042)</td>
<td>(0.046)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ 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\textsuperscript{a,b}

<table>
<thead>
<tr>
<th>Type of Crime</th>
<th>All Hours and Periods</th>
<th>Daytime\textsuperscript{c}</th>
<th>Nighttime\textsuperscript{c}</th>
<th>COVID-19 Lockdown\textsuperscript{d}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Retail Employment</td>
<td>% Retail Employment</td>
<td>% Retail Employment</td>
<td>% Retail Employment</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Property crime</td>
<td>2.307***</td>
<td>-0.085***</td>
<td>2.650***</td>
<td>1.685***</td>
</tr>
<tr>
<td></td>
<td>(0.165)</td>
<td>(0.015)</td>
<td>(0.182)</td>
<td>(0.198)</td>
</tr>
<tr>
<td>Grand Larceny &amp; Burglary</td>
<td>1.792***</td>
<td>-0.053***</td>
<td>2.034***</td>
<td>1.766***</td>
</tr>
<tr>
<td></td>
<td>(0.159)</td>
<td>(0.014)</td>
<td>(0.179)</td>
<td>(0.223)</td>
</tr>
<tr>
<td>Petit Larceny</td>
<td>2.650***</td>
<td>-0.104***</td>
<td>3.050***</td>
<td>1.733***</td>
</tr>
<tr>
<td></td>
<td>(0.189)</td>
<td>(0.018)</td>
<td>(0.208)</td>
<td>(0.238)</td>
</tr>
<tr>
<td>Theft of Services &amp; Fraud</td>
<td>1.819***</td>
<td>-0.087***</td>
<td>2.204***</td>
<td>1.298***</td>
</tr>
<tr>
<td></td>
<td>(0.319)</td>
<td>(0.030)</td>
<td>(0.404)</td>
<td>(0.459)</td>
</tr>
<tr>
<td>Robbery</td>
<td>2.330***</td>
<td>-0.144***</td>
<td>2.217***</td>
<td>2.265***</td>
</tr>
<tr>
<td></td>
<td>(0.220)</td>
<td>(0.021)</td>
<td>(0.254)</td>
<td>(0.265)</td>
</tr>
<tr>
<td>Auto Theft</td>
<td>0.833***</td>
<td>-0.057**</td>
<td>0.586*</td>
<td>1.152***</td>
</tr>
<tr>
<td></td>
<td>(0.212)</td>
<td>(0.023)</td>
<td>(0.333)</td>
<td>(0.320)</td>
</tr>
</tbody>
</table>

\textsuperscript{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).

\textsuperscript{b} Every pair of coefficients, share of retail employment and spatial concentration of retail, is obtained from the preferred specification in Table 2 column (4). Note that the coefficients for property crime during all hours are identical to those in Table 2 column (4). The sample for all models is 3,507 grid cells (0.2 square miles) in the NYC area. Crime data refers to incidents in 2018.

\textsuperscript{c} Daytime hours include crimes between 10 am and 6 pm. Nighttime hours include crimes between 10 pm and 5 am.

\textsuperscript{d} COVID-19 lockdown refers to the first two weeks of the NYC lockdown, March 22\textsuperscript{nd} to April 5\textsuperscript{th} of 2020.
Table 4: Alternate Measures of Spatial Concentration

Panel A: Correlation between alternate measures of retail spatial concentration

<table>
<thead>
<tr>
<th></th>
<th>$\tilde{G}_{Ret,Emp}$</th>
<th>$\tilde{G}_{Ret,Stores}$</th>
<th>$\tilde{G}_{Ret,Sales}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{G}_{Ret,Emp}$</td>
<td>1.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\tilde{G}_{Ret,Stores}$</td>
<td>0.15</td>
<td>1.0</td>
<td>-</td>
</tr>
<tr>
<td>$\tilde{G}_{Ret,Sales}$</td>
<td>0.54</td>
<td>0.14</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Panel B: Crime and police stop regressions

<table>
<thead>
<tr>
<th>Regression Sample</th>
<th>Property Crime</th>
<th>Police Stops</th>
<th>Property Crime</th>
<th>Police Stops</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Retail Employment</td>
<td>$\tilde{G}_{Ret,Emp}$</td>
<td>$\tilde{G}_{Ret,Stores}$</td>
<td>$\tilde{G}_{Ret,Sales}$</td>
</tr>
<tr>
<td>All Hours/Periods</td>
<td>2.646***</td>
<td>-0.084***</td>
<td>-0.122***</td>
<td>0.002</td>
</tr>
<tr>
<td>Daytime</td>
<td>2.974***</td>
<td>-0.101***</td>
<td>-0.120***</td>
<td>-0.001</td>
</tr>
<tr>
<td>Nighttime</td>
<td>2.109***</td>
<td>-0.021</td>
<td>-0.141***</td>
<td>0.028</td>
</tr>
<tr>
<td>COVID-19 Lockdown</td>
<td>3.096***</td>
<td>-0.078**</td>
<td>-0.142***</td>
<td>0.052</td>
</tr>
</tbody>
</table>

* 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. 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 Police stops include discretionary stops pooled from 2016-2018. During COVID-19 lockdown police stops were likely motivated by other confounding factors such as compliance of the stay at home executive orders, for that reason results for police stops during lockdown are not estimated.
d Spatial concentration of retail employment, $\tilde{G}_{Ret,Emp}$, 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, $\tilde{G}_{Ret,Stores}=\sum_{e} \omega_{le}(d_{ie}) / n_{l}(d_{le})$, where $n_{l}$ is the number of nearby establishments within a given distance, $d_{ie}$, and $\omega_{le}$ is defined as in (3.2).
e Daytime hours include crimes between 10 am and 6 pm. Nighttime hours include crimes between 10 pm and 5 am.
f COVID-19 lockdown refers to the first two weeks of the NYC lockdown, which refers to March 22nd to April 5th of 2020.
<table>
<thead>
<tr>
<th>Panel A: Retailer cost of space/$ sold</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Workers/Sales</td>
<td>71,574***</td>
<td>70,893***</td>
<td>71,129***</td>
<td>71,828***</td>
<td>37,983***</td>
</tr>
<tr>
<td></td>
<td>(6,196)</td>
<td>(6,116)</td>
<td>(6,116)</td>
<td>(6,432)</td>
<td>(7,180)</td>
</tr>
<tr>
<td>Spatial Concentration of Retail: $\hat{g}_{Ret,Emp}$</td>
<td>0.373***</td>
<td>0.413***</td>
<td>0.436***</td>
<td>0.384***</td>
<td>0.353***</td>
</tr>
<tr>
<td></td>
<td>(0.106)</td>
<td>(0.116)</td>
<td>(0.115)</td>
<td>(0.123)</td>
<td>(0.120)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,596</td>
<td>1,596</td>
<td>1,596</td>
<td>1,596</td>
<td>1,596</td>
</tr>
<tr>
<td>R2</td>
<td>0.47</td>
<td>0.48</td>
<td>0.48</td>
<td>0.50</td>
<td>0.52</td>
</tr>
</tbody>
</table>

| Panel B: Retailer non-inventory cost/$ sold | | | | | |
| Spatial Concentration of Retail: $\hat{g}_{Ret,Emp}$ | 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: $\hat{g}_{Ret,Emp}$ | 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 $\hat{g}_{Ret,Emp}$ | 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: Data Sources, Access, and Variable Construction

A.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. 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.

A.2 Data sources and access

A.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. 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 data source used to create each measure in the paper.

A.2.2 Data Sources


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


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.
A.2.3 Data sources for each measure in the paper

Table A-1 below provides a complete listing of each variable and its source used in different parts of the paper.
<table>
<thead>
<tr>
<th>Variable Groups</th>
<th>Variable</th>
<th>Definition</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variables</td>
<td>Neighborhood property crimes</td>
<td>Total number of property crimes occurred in the grid square. Property crime refers to burglary, grand larceny, petit larceny, theft of services or fraud.</td>
<td>New York Police Department (NYPD)</td>
</tr>
<tr>
<td></td>
<td>Neighborhood police stops</td>
<td>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.</td>
<td>New York Police Department (NYPD)</td>
</tr>
<tr>
<td></td>
<td>Establishment input cost per dollar sold</td>
<td>Total expenditures on space and labor divided by the level of sales at the establishment level. See expression (2.14).</td>
<td>Dun &amp; Bradstreet and CompStak Inc.</td>
</tr>
<tr>
<td>Baseline Controls</td>
<td>Distribution and level of employment</td>
<td>Total employment and employment in each selected industry (SIC)</td>
<td>Dun &amp; Bradstreet</td>
</tr>
<tr>
<td></td>
<td>Spatial Concentration Measures ($G$)</td>
<td>Spatial $G$ for selected industries based on employment and sales</td>
<td>Dun &amp; Bradstreet</td>
</tr>
<tr>
<td></td>
<td>Monthly visits to POI</td>
<td>Average monthly visits to each POI in a grid square based on cellphone data</td>
<td>SafeGraph</td>
</tr>
<tr>
<td></td>
<td>Number of Trees</td>
<td>Total number of trees on the street based on 2015 Street Tree Census</td>
<td>Department of Parks and Recreation</td>
</tr>
<tr>
<td></td>
<td>Average Building Age</td>
<td>Average age of buildings across the grid square.</td>
<td>MapPLUTO</td>
</tr>
<tr>
<td></td>
<td>Average Building Assessed Value</td>
<td>Average assessed value of the building in the grid square</td>
<td>MapPLUTO</td>
</tr>
<tr>
<td></td>
<td>Overlapping Police Precincts</td>
<td>Grid square overlaps multiple police precincts</td>
<td>MapPLUTO</td>
</tr>
<tr>
<td></td>
<td>Share of Residential Units Within Bldgs</td>
<td>Share of all units in the grid square that are residential</td>
<td>MapPLUTO</td>
</tr>
<tr>
<td></td>
<td>Neighborhood Sales per Worker</td>
<td>Total sales of single-site establishments in the grid square over total employment</td>
<td>MapPLUTO</td>
</tr>
<tr>
<td>Establishment Characteristics</td>
<td>Neighborhood Average Market Risk</td>
<td>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</td>
<td>Dun &amp; Bradstreet</td>
</tr>
<tr>
<td></td>
<td>Neighborhood Avg. Establishment Age</td>
<td>Establishment Age = 2019 – Founding Year</td>
<td>Dun &amp; Bradstreet</td>
</tr>
<tr>
<td>Zoning Restrictions</td>
<td>Share Special District</td>
<td>Share of lots in the grid located in special purpose districts.</td>
<td>MapPLUTO</td>
</tr>
<tr>
<td></td>
<td>Share Commercial allowed in Residential</td>
<td>Share of lots in the grid that are allow for commercial overlay within a residential zoning district</td>
<td>MapPLUTO</td>
</tr>
<tr>
<td></td>
<td>Share Multiple Zoning</td>
<td>Share of lots in the grid that are between multiple zoning features.</td>
<td>MapPLUTO</td>
</tr>
<tr>
<td></td>
<td>Average Density Allowed by Residential Zoning</td>
<td>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.</td>
<td>MapPLUTO</td>
</tr>
<tr>
<td></td>
<td>Average Density Allowed by Commercial Zoning</td>
<td>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.</td>
<td>MapPLUTO</td>
</tr>
<tr>
<td></td>
<td>Share buildings with height restriction</td>
<td>Share of lots in the grid that are in a limited height district</td>
<td>MapPLUTO</td>
</tr>
<tr>
<td></td>
<td>Average residential FAR</td>
<td>Maximum allowable residential floor area ratio across the grid</td>
<td>MapPLUTO</td>
</tr>
<tr>
<td></td>
<td>Average commercial FAR</td>
<td>Maximum allowable commercial floor area ratio across the grid</td>
<td>MapPLUTO</td>
</tr>
<tr>
<td>Distance to Landmarks</td>
<td>Distance Central Park</td>
<td>Distance between the grid centroid and Central Park</td>
<td>Department of Parks and Recreation</td>
</tr>
<tr>
<td></td>
<td>Distance Nearest Park</td>
<td>Distance from grid centroid to nearest park</td>
<td>Department of Parks and Recreation</td>
</tr>
<tr>
<td></td>
<td>Distance Nearest Subway</td>
<td>Distance from grid centroid to nearest subway entrance (0 if entrance inside grid)</td>
<td>Metropolitan Transportation Authority</td>
</tr>
<tr>
<td></td>
<td># Subway Entrances</td>
<td>Number of subway entrances inside the grid square</td>
<td>Metropolitan Transportation Authority</td>
</tr>
<tr>
<td>Variable Groups</td>
<td>Variable</td>
<td>Definition</td>
<td>Source</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Amenities</td>
<td>Average PM 2.5</td>
<td>Annual average fine particulate matter &lt; 2.5 microns (2018), 300 mt resolution</td>
<td>Community Air Survey Air Pollution</td>
</tr>
<tr>
<td></td>
<td>Ln(Reported rat problems)</td>
<td>Total 2018 rodent inspections that resulted in active rat signs.</td>
<td>Department of Health and Mental Hygiene</td>
</tr>
<tr>
<td></td>
<td>Ln(Failed rodent inspections)</td>
<td>Total 2018 rodent inspections that did not pass the inspection.</td>
<td>Department of Health and Mental Hygiene</td>
</tr>
<tr>
<td></td>
<td>Ln(Complaints about traffic lights)</td>
<td>311 Complaints (requests) related to traffic signal condition</td>
<td>Department of Information Technology and Telecommunications (DITT)</td>
</tr>
<tr>
<td></td>
<td>Ln(Complaints about street lights)</td>
<td>311 Complaints (requests) related to street light condition</td>
<td>DITT</td>
</tr>
<tr>
<td></td>
<td>Newly planted trees 2005-2015</td>
<td>Difference between 2005 and 2015 Tree Census</td>
<td>Department of Parks and Recreation</td>
</tr>
<tr>
<td>Historic Places and Landmarks</td>
<td>Historic places registered before 2018 to the National Register of Historic Places</td>
<td></td>
<td>NY State Parks, Recreation and Historic Preservation</td>
</tr>
<tr>
<td>Public alarm boxes on the street</td>
<td>Fire alarm boxes in the grid: includes Emergency Reporting System (ERS) and Box Alarm Reporting System (BARS)</td>
<td></td>
<td>Fire Department</td>
</tr>
<tr>
<td>Active Sites: HIV testing and condom distribution locations</td>
<td>Active venues distributing free safer sex products under the NYC Condom Availability Program – HIV.</td>
<td></td>
<td>Department of Health and Mental Hygiene</td>
</tr>
<tr>
<td>Building with Irregular shape and Tax exemptions</td>
<td>Share of buildings with irregular shape</td>
<td>Share of lots in the grid that have an irregular shape</td>
<td>MapPLUTO</td>
</tr>
<tr>
<td></td>
<td>Share of buildings that are tax exempt</td>
<td>Share of lots in the grid that have at least 20% of their assessment value exempt of property tax.</td>
<td>MapPLUTO</td>
</tr>
</tbody>
</table>
Appendix B: Robustness Checks

Table B-1 summarizes estimates for three robustness checks for the more fully specified crime and police stop models in columns 4 and 8 of Table 2. The Table 2 estimates are repeated in columns 1 (for crime) and 5 (for police stops) to facilitate comparison. Columns 2 and 6 add 25 additional neighborhood controls to the models. Columns 3 and 7 measure $\tilde{G}_{Emp}$ using all single-site establishments and all establishments that belong to multi-site firms.

The models in columns 4 and 8 use a different neighborhood design that requires a brief description. In these models, each neighborhood is 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:

$$G_i = \sum_{j=1}^{9} \left( \frac{E_{ij}}{\sum_{j=1}^{9} E_{ij}} \right)^2$$

(B.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. As a further robustness check, we used the same design as above but restricted our estimating sample to non-overlapping neighborhoods. Results were quite similar to those in Table B1 (columns 4 and 8).

The important take away from Table B-1 is that the coefficients on spatial concentration are similar to the core models from Table 2. This includes that spatial concentration of retail employment deters crime and reduces police stops. It also includes that spatial concentration of retail employment has a more notable effect on crime and police stops than concentration of employment for the other industries highlighted in the table (finance, manufacturing and service).

---

41 Sample size for the 9-grid square neighborhoods is larger than for the individual 0.2 square mile neighborhoods used in the main body of the paper. This is because the larger area used to define a neighborhood (nine grid squares) 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.
Table B-1: Alternate Estimates of Columns 4 and 8 of Table 2

| Property Crime | | | | | | | | Police Stops | | | | | |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|                | Table 2 Column 4 | 25 Added Neigh Controls | Use Multi-Site Firms for $\hat{g}_{Emp}$ | 9 Square Neigh Designc | Table 2 Column 8 | 25 Added Neigh Controls | Use Multi-Site Firms for $\hat{g}_{Emp}$ | 9 Square Neigh Designc | | | | |
| (1)            | (2)            | (3)            | (4)            |                  | (5)            | (6)            | (7)            | (8)            | | | | |
| Log total emp  | 0.554***       | 0.374***       | 0.547***       | 0.665***        | 0.349***       | 0.063          | 0.341***       | 0.378***       | | | | |
|                | (0.022)        | (0.025)        | (0.023)        | (0.014)         | (0.056)        | (0.060)        | (0.056)        | (0.036)        | | | | |
| Share emp: Retail | 2.307***      | 2.080***       | 2.247***       | 2.879***        | 2.017***       | 0.762**        | 2.319***       | 1.643***       | | | | |
|                | (0.165)        | (0.167)        | (0.166)        | (0.163)         | (0.359)        | (0.358)        | (0.362)        | (0.388)        | | | | |
| Share emp: Fin | -0.451**       | 0.001          | -0.577**       | -1.002***       | -0.752         | -0.461         | -0.217         | -3.230***       | | | | |
|                | (0.217)        | (0.213)        | (0.226)        | (0.204)         | (0.608)        | (0.537)        | (0.625)        | (0.556)        | | | | |
| Share emp: Manf | -0.226         | -0.171         | -0.108         | 0.412           | 0.027          | 0.431          | 0.325          | -0.650          | | | | |
|                | (0.247)        | (0.237)        | (0.260)        | (0.321)         | (0.536)        | (0.560)        | (0.546)        | (0.637)        | | | | |
| Share emp: Serv | -0.025         | 0.079          | -0.044         | 0.363***        | 0.055          | -0.056         | -0.183         | -0.101          | | | | |
|                | (0.129)        | (0.112)        | (0.129)        | (0.114)         | (0.318)        | (0.316)        | (0.319)        | (0.269)        | | | | |
| Log POI visits | 0.516***       | 0.299***       | 0.526***       | 0.611***        | 0.894***       | 0.537***       | 0.900***       | 1.284***        | | | | |
|                | (0.030)        | (0.029)        | (0.030)        | (0.023)         | (0.069)        | (0.068)        | (0.071)        | (0.055)         | | | | |
| $\tilde{g}_{Ret,Emp}$ | -0.085***     | -0.073***      | -0.056***      | -1.058***       | -0.110***      | -0.067**       | -0.223***      | -2.235***        | | | | |
|                | (0.015)        | (0.014)        | (0.016)        | (0.125)         | (0.036)        | (0.029)        | (0.044)        | (0.322)         | | | | |
| $\tilde{g}_{Fin,Emp}$ | 0.005          | 0.014          | 0.017          | -0.179          | 0.042          | 0.04           | -0.048         | 0.693***         | | | | |
|                | (0.016)        | (0.015)        | (0.018)        | (0.112)         | (0.040)        | (0.033)        | (0.039)        | (0.234)         | | | | |
| $\tilde{g}_{Manf,Emp}$ | 0.007          | -0.001         | -0.012         | 0.036           | -0.020         | 0.002          | -0.071         | -0.357***        | | | | |
|                | (0.018)        | (0.017)        | (0.019)        | (0.062)         | (0.041)        | (0.039)        | (0.045)        | (0.132)         | | | | |
| $\tilde{g}_{Serv,Emp}$ | -0.029         | 0.035*         | -0.012         | -0.647***       | -0.137***      | 0.038          | 0.057          | -1.120***        | | | | |
|                | (0.019)        | (0.020)        | (0.020)        | (0.145)         | (0.046)        | (0.044)        | (0.046)        | (0.248)         | | | | |
| Other Neigh Var | 6              | 31             | 6              | 6               | 6              | 31             | 6              | 6               | | | | |
| Overdispersion  | 0.42           | 0.32           | 0.421          | 0.217           | 2.03           | 1.50           | 2.004          | 1.115           | | | | |
| Observations   | 3,506          | 3,504          | 3,546          | 5,122           | 3,506          | 3,504          | 3,546          | 5,122           | | | | |

* 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 unit of analysis is 0.6 miles overlapping grid squares, defined as the usual target 0.2 mi grid plus its eight neighboring cells.

b Only self-initiated police stops between 2016-2018 are included. Stops initiated by 911 calls and ongoing investigations are removed.

c 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. For columns (2) and (6) in which 25 more controls are included, those are listed in Table A1 and refer to Establishment Characteristics, Zoning Restrictions, Distance to Landmarks, Amenities, and Building with Irregular shape and Tax exemptions.