The Determinants of Criminal Victimization in São Paulo State, Brazil*

Fábio Augusto Reis Gomes**
Lourenço Senne Paz***

Abstract

In this paper, the determinants of victimization by burglary/larceny are assessed theoretically by a model that takes into account the victim's investment in self-protection and criminals with heterogeneous ability. Such features generate a non-increasing relationship between an individual's income (which is attractive to the criminal) and her victimization likelihood. Thus, we obtain a novel result indicating that rich people might experience less victimization than middle-class individuals. The determinants are also investigated empirically using a victimization survey for São Paulo state – Brazil (*Pesquisa de Condição de Vida – 1998*), and unlike the previous literature, we found evidence that most of the city size effect on victimization likelihood is explained by the city's population characteristics (income, education, etc.). Moreover, crime and city size do not have an increasing relationship, and the same applies to victimization likelihood and victim's income. These results are in line with our theoretical model predictions. Finally, we found evidence that people do not consider their victimization likelihood when choosing a city to live in.

Keywords: Victimization, Burglary/Larceny, City Size, São Paulo State, Brazil.

JEL Codes: K40, K42, J28.

---

*Submitted in February 2007. Revised in March 2009. We would like to thank Luiz Renato Lima, Luis Braido, Cleomar Gomes, the Editor and three anonymous referees for important comments and suggestions that significantly improved this paper. The usual disclaimer applies. Fábio Augusto Reis Gomes thanks CNPq for financial support.

**Insper Instituto de Ensino e Pesquisa, Rua Quatá, 300, sala 422, Vila Olímpia. São Paulo – SP, Brasil. E-mail: fabioarg@isp.edu.br

***University of Maryland – Dept. of Economics, 3105 Tydings Hall College Park- MD, 20742-7211, USA. E-mail: pazlourenco@yahoo.com

1. Introduction

Developing countries have experienced an increase in crime rates recently. In some places, such as São Paulo state in Brazil, the crime rate is so high that at least one in every eight families had a member who fell victim to burglary or larceny in 1998.[1] Moreover, Khan (1999) estimates a lower bound of the costs imposed by high crime rates in São Paulo state to be 3% of its gross domestic product. From an economic perspective, this is truly an important topic.

Lack of detailed individual-level victimization data has been a major hindrance to the economics of crime literature. Nevertheless, in some cases, the few surveys available can be used to confront the predicted signs from theoretical models with the estimated signs from reduced-form specifications. Two important papers in this fashion are Glaeser and Sacerdote (1996) (hereafter GS) using U.S. data, and Gaviria and Pagés (2002) (hereafter GP) using data for Latin America. They uncover some interesting findings. The first is that crime rate increases with city size. Second, individual characteristics account for a small part of the city size effect on crime. Third, recent city growth is positively correlated with crime rate. Fourth, victimization likelihood also increases with the individual income.

The current literature has ignored two important aspects. The first is that potential victims can invest in self-security. From a theoretical perspective, this kind of investment would be able to generate non-monotonic relationships between victimization and city size, and victimization and income in the case of criminals with different skill levels. The other aspect is the possibility of self-selection into city size, i.e., individuals prone to be victimized self-select themselves to cities of a certain size. Therefore, city size would be mistakenly interpreted as the main determinant of victimization, and estimations of city size effect will not be consistent.

Our paper will deal with these two aspects. First, we augment the GP model in order to discuss the role of investment in self-protection when criminals have heterogeneous ability. Indeed, we show that if the investment is concave in income, in larger cities richer individuals will invest in security, counterbalancing income attractiveness, and that is how the non-monotonic relationships between crime and city size and between crime and income are generated.

Second, using the victimization data from SEADE's 1998 Living Condition Survey (*Pesquisa de Condição de Vida*, hereafter PCV) we first examine if self-selection into city size is indeed relevant to the individual victimization problem. The results did not support the self-selection issue. Thus, following the GS and GP methodology, we verify if the four stylized facts listed above are present in our data. To the best of our knowledge, this is the first paper to use this dataset to investigate the determinants of victimization.[2]

[1] See Table 2.

[2] Kilsztjn et al. (2003), SEADE (1999) and Catão (2000) analyzed the same dataset using

Our results indicate that city size is still an important factor in the likelihood of becoming a victim of burglary/larceny, and crime rate does not increase with city size anymore. As a matter of fact, medium-sized cities (500,000 to 1 million inhabitants) have higher crime rates than large cities (above 1 million inhabitants). Moreover, contrary to GP and GS findings, the potential victim characteristics such as income, education, etc. account for a large part of the city size effect, and the recent city population growth seems to have no role at all in explaining the likelihood of being a victim. Last, victimization likelihood does not increase with the victim's income.

This paper is organized as follows. Section 2 presents a review of the literature and our theoretical model. The dataset is described in Section 3. Section 4 depicts the estimated models and reports the results. Finally, the conclusions are drawn in Section 5.

2. The Determinants of Property Crime Victimization

2.1 Literature review

Crime has been studied in several fields of Social Sciences. In the sociology literature,[3] an important and well known approach is the opportunity model. This model focuses on the occurrence of crime that happens if and only if a criminal meets a potential victim under favorable conditions. As a result, the factors driving these meetings and the favorable conditions are the variables of interest (Newman (1972), Reppeto (1974)).

Further steps in this approach were given by Hindelang et al. (1978) and Cohen and Felson (1979). Hindelang et al. (1978) allowed for the potential victim's lifestyle (the way people allocate their time between work and leisure) to impact the chance of becoming a victim, by affecting their proximity to criminals.[4] Cohen and Felson (1979) developed the routine activity approach, which also takes into account the individual's allocation of time to explain how the potential victim meets a criminal in a situation in which there is not enough security to prevent the crime.

While Bursik and Grasmick (1993) emphasized the proximity between both types of agents as a crucial determinant of victimization, Clarke and Felson (1993) argued that the security level plays a central role. Indeed, Cohen and Felson (1979), and Felson and Cohen (1980) discussed that the victim's security depends on the presence of persons or objects that can prevent the criminal activity. However, as pointed out by Jensen and Brownfield (1986), a major drawback of both models is that they do not model the criminal's decision, in particular her time allocation.

descriptive statistics.

[3]For a more comprehensive review of this literature, see Beato et al. (2004) and Peixoto et al. (2007).

[4]See also Gottfredson and Hindelang (1981) and Garafalo (1987).

Cohen et al. (1981) extended the original routine activity model, including two new variables: exposure and attractiveness of potential victims. Exposure refers to the physical visibility and accessibility of persons or objects to potential offenders at any time or place; while target attractiveness means the material or symbolic desirability of persons or property targets to potential offenders.

In the economics literature, the seminal paper of GS analyzed a victimization survey conducted in several U.S. cities. Their most important finding is that there is more crime in cities, especially large cities, in relation to rural areas, even after controlling for dwellers' characteristics. The paper by GP provided a theoretical model in which they incorporated the insights of the sociological approach into an economic framework. We will discuss their model in more detail and extend it in the next section. GP also empirically analyzed the determinants of property crime victimization in Latin American cities, concluding that the typical victim of crime in Latin America comes from rich and middle-class households and tends to live in larger cities. They conjecture that this effect comes from the lower arrest probability in large cities and from the fact that larger cities harbor a greater proportion of crime-prone individuals.

Fajnzylber et al. (2001) presented a very good review of other papers about victimization. Among them, Piquet and Fajnzylber (2001) is one of the first to study victimization issues in Brazil. This paper used the 1996 FGV-ISER Rio de Janeiro city victimization survey and the 1988 PNAD victimization supplement for São Paulo city and investigated the individual characteristics effect on victimization likelihood. Another interesting paper is Beato et al. (2004), which used CRISP (2002) Belo Horizonte victimization survey and estimated the effect of individual lifestyle and housing characteristics on victimization. Nevertheless, the data from the last two papers did not allow the researchers to verify the city size effect on crime, as done in GS and GP, and as we will do in this paper.

2.2 Theoretical model

Suppose that a city has N risk-neutral individuals, in which n_c are criminals and $n_v = N - n_c$ are potential victims. The attractiveness is the individual's income, w_i, and the offender's j benefit of a crime is given by $\delta_j w_i$, in which $\delta_j \sim U[\delta_L, \delta_H]$ is an offender specific parameter reflecting either her ability or preference for certain types of crime. This is the first departure in relation to GP, which considers only one type of criminal. If caught by the police, the punishment received is $F_j = F(\delta_j)$, where $F'(\delta_j) > 0$ and $F_j \sim U[F_L, F_H]$. The potential victim can invest e_i in her self-security taking into account the distribution of δ_j and F_j. Victim's characteristics, w_i and e_i, are observable to criminals.

The offender has a $1 - p(e_i)$ likelihood of having a successful offense against victim (i), which is decreasing in e_i, i.e. $p'(\bullet) > 0$. Thus, when the offender meets the potential victim, a crime is committed if its expected benefit, $[1 - p(e_i)]\delta_j w_i$, surpasses its expected punishment, $p(e_i)F_j$. And, given their risk neutrality, potential victims consider only the average criminal which has ability $\triangle_\delta \equiv E_i(\delta_j) = (\delta_H + \delta_L)/2$ and punishment cost $\triangle_F \equiv E_i(F_j) = (F_H + F_L)/2$.[5] Therefore, the chosen e_i level will be the one that makes the average offender indifferent about committing or not the crime, and it is given by $e_i^* = e_i^*(w_i)$:

$$[1 - p(e_i^*)] \triangle_\delta w_i = p(e_i^*)\triangle_F \Rightarrow e_i^* = p^{(-1)} \left[\frac{\triangle_\delta w_i}{\triangle_F + \triangle_\delta w_i} \right] \tag{1}$$

But, this investment in security (*guardianship*) takes place if and only if it does not exceed the crime expected loss, $E(\delta_j w_i) = \triangle_\delta w_i$.[6] The victim meets the criminal with probability $\theta = n_c/(N - 1)$, as a result, the expected loss for the victim for every meeting is $\theta \triangle_\delta w_i$. Moreover, suppose that a victim meets a certain number of persons, a fraction σ of the society with reposition, then the total expected loss becomes $\sigma N \theta \triangle_\delta w_i$. Hence, potential victims invest in private protection if $e_i^* < \sigma N \theta \triangle_\delta w_i$. The expected loss is increasing in the number of city's inhabitants (N), a measure of exposure (σ), an approximation of the share of criminals in the population (θ), victim's income (w_i) and mean criminal ability $(\triangle_\delta)$.

Suppose that $e_i^*(w_i)$ is concave, i.e., the security investment cost is decreasing. Then, only those agents who have $w_i > w_i^*$ will invest in security, as shown in Figure 1.[7]

[5]Investment in self-protection can not completely prevent crime in our model, which is a realistic scenario not contemplated by the GP model.

[6]We are considering the interesting case in which the average criminal's ability is high enough such that $[1 - p(0)]\triangle_\delta w_i > p(0)\triangle_F$ for all w_i. In the other case, the expected loss is zero and nobody invests in self-protection, which is rather unrealistic.

[7]When $e_i^*(w_i)$ is convex, only those agents whose $w_i > w_i^*$ will not invest in self-security, which is at odds with empirical evidence, for example Beato et al. (2004).

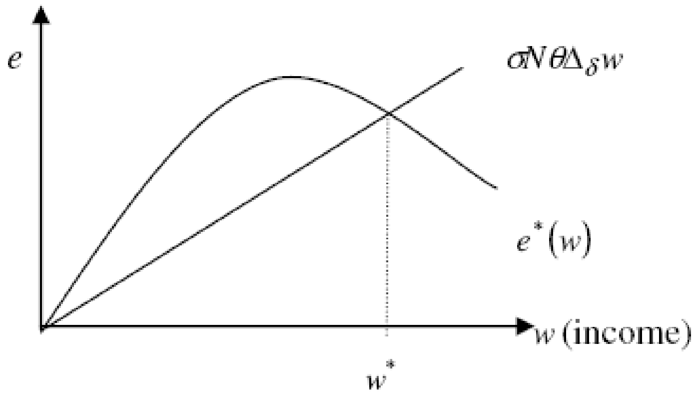

Figure 1
$e_i^*(w_i)$ (cost of investing in private security) is concave

Thus, the agent chooses either $e_i^* = 0$ or $e_i^* = p^{(-1)}[\triangle_\delta w_i/(\triangle_F + \triangle_\delta w_i)]$, according to her wealth level. Since the potential targets meet criminals with probability θ, the likelihood of an agent i being attacked by a criminal j is given by V_{ij}:

$$V_{ij} = \theta \times Pr\left[p(e_i^*) < \frac{\delta_j w_i}{F_j + \delta_j w_i}\right] \tag{2}$$

Because $p'(\bullet) > 0$, V_{ij} is lower when agent i invests in self-security, as predicted by the previously discussed models. And, as punishment gets harsher, V_{ij} becomes zero eventually, and there will be no crime in accordance with Becker's (1968) deterrence hypothesis.

Notice that, *ceteris paribus*, we expect that in larger cities a larger number of persons will invest in private protection, since the total expected loss increases with city size. Figure 2 illustrates this case, where $N_1 > N_2$. Individuals with income between w_1 and w_2 will invest in private security if they live in a larger city, due to the economy of scale in self-protection investments, e.g. by building a wall, an individual would be protected against more criminals than if she builds the wall in a smaller city. Since the cost of investment is independent of city size, the potential victim has more to lose in large cities, and therefore she will invest more in self-protection. This effect is stressed if:

i) Larger cities harbor a greater proportion of crime-prone individuals, $\theta(N_1) > \theta(N_2)$;

ii) There exists positive externality or learning among criminals, $\triangle_\delta(N_1) > \triangle_\delta(N_2)$.

In other words, the larger the city, the larger the incentives to invest in self-protection.

Notice that the cost of not investing in self-protection is larger in big cities; even in the case in which all cities have the same share of criminals (θ is fixed). Hence, individuals with the same income (and other observable characteristics) living in cities with different population will present a different behavior toward self-protection and will have a different victimization likelihood. In this case, we expect that individuals in the middle-income range would be more victimized in smaller cities (see Figure 2).

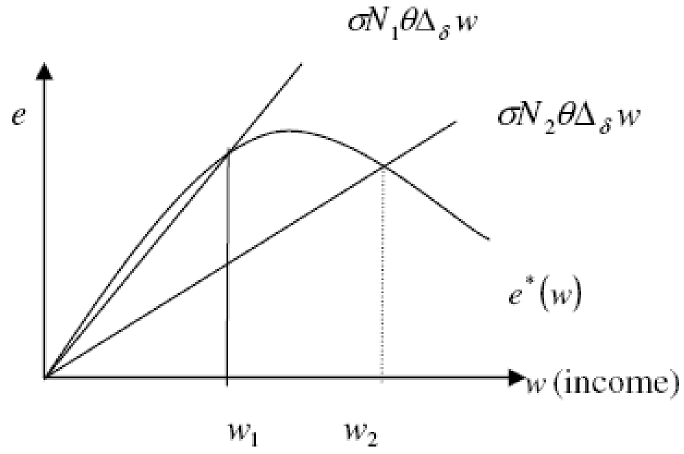

Figure 2
Different city sizes and investment in private security (e), for $N_1 > N_2$

3. Data

A major empirical issue in victimization studies is the lack of reliability of official data, in the sense that not all crimes committed are reported to the police. This underreporting bias can be really large, and among its causes we can cite the victim's assessment of police ability (Catão, 2000), lack of interest by the victim, the victim may be ashamed of disclosing sexual offenses, and the victim might not recognize the occurrence as a crime (e.g. fight among classmates).

A partial solution to this problem, in particular to the first cause, is the use of victimization surveys, adopted by GS, GP, and us. These victimization surveys have been conducted in the USA since the 1970s and for some Latin American countries since the 1990s. A problem inherent to any survey is also present here: the respondent might recall only the most recent or serious event. Even after all these pitfalls, Soares (2004) provided evidence that victimization surveys present higher crime rates than those reported in police records, and they seem to be the most appropriate data available.

In Brazil, there are some victimization surveys covering some cities, or at most a single state. Unfortunately, they are not frequent and use different methodologies, making it hard to draw any comparison across them. The victimization survey used in this paper is the SEADE's 1998 Living Condition Survey (*Pesquisa de Condição de Vida*). It covers São Paulo state, which has about 37 million inhabitants spread into 643 cities, and produces about one third of the Brazilian GDP. The PCV is a household survey that contains information on demographics, housing, employment, income, and exposure to violence of all household members.

The survey was conducted between June and November 1998 in seven regions of São Paulo state, encompassing a total of 15,000 domiciles in 104 cities, all of which have more than 50,000 inhabitants. Due to missing data, we were able to use information from 9,354 domiciles, which accounted for 9,849 families and 28,853 individuals.

Exposure to violence is assessed by asking whether the individual was a victim of burglary or larceny[8] in the 12 months preceding the interview. The survey has no information about where the offense happened. We assumed it happened in the victim's city of residence.

The data about city populations and population growth rates were obtained from SEADE (2004). The annual population growth rate was calculated as the annual rate between 1998 and the previous one, two, five and 10 years. These growth rates presented a positive correlation of 0.95. So, we chose to use the two-year growth, i.e., the annual population growth rate between 1996 and 1998. It is worth noting that results are robust to any of those growth rates.

In Table 1, the variables used are presented with their respective unconditional means and standard deviations. We can see that 40% of São Paulo state inhabitants live in cities with more than one million people and less than 10% live in cities with less than 100,000 inhabitants. Interestingly, about six percent of the whole São Paulo state inhabitants were victims of burglary/larceny, which is a very high crime rate.

The figures are more dramatic if we consider crime incidence per family, as shown in Table 2. In the central region, one in every eight families had a member

[8]The PCV considers burglary and larceny as equivalent crimes, even though burglary is a crime without physical violence while larceny involves physical violence threat and is considered more serious.

Table 1
Descriptive statistics of the main variables

Variable	Mean	Std. deviation
Burglary/larceny victims (% of population)	5.81	
Men (%)	48.13	
Black/Other (%)	27.77	
Asian (%)	1.33	
Migrant (%)	48.17	
Employed, retired or studying (%)	74.39	
Living in cities with less than 100,000 inhabitants (%)	9.689	
Living in cities with 100,000 to 500,000 inhabitants (%)	37.59	
Living in cities with 500,000 to 1,000,000 inhabitants (%)	11.44	
Living in cities with more than 1,000,000 inhabitants (%)	41.28	
Living in metropolitan Areas (%)	65.99	
Household that owns a car (%)	48.89	
Marginal household in (%)	21.71	
Age (years)	29.73	18.72
Years living in current home	18.91	16.01
Years of education	6.49	4.24
Household total income (local currency)	1522.55	1531.05
Household per capita income (local currency)	426.83	475.72
Number of persons per home	4.23	1.77
Annual population growth in the last two years (%)	4.92	1.99

victimized, and in São Paulo metropolitan area the number increases to one in every five. As GS first noted for U.S. data, crime incidence tends to be larger in metropolitan areas (MA) and in large cities, and the same pattern emerges in Table 2 for the unconditional victimization rates.

Table 3 reports the percentage of population that was victim of burglary/larceny according to city size ranges. The four ranges were chosen in a similar fashion to GS and GP. The first range includes cities with more than 50,000 and less than 100,000 inhabitants. The next range is for cities with less than 500,000, but more than 100,000 people. The third range encompasses cities with more than 500,000 and less than one million inhabitants. The last range includes cities with more than one million people.

Table 2
Victimization by São Paulo state regions

Region	São Paulo MA	Central	East	Santos MA	North	West	Vale do Paraíba
Burglary/Larceny							
Individuals (%)	6.6	4.1	4.7	6.8	4.5	4.2	5.6
Family (%)	19.9	12.2	13.9	19.3	14.4	12.3	17.1

Note: Family means at least one person of the family. MA refers to metropolitan area.

Table 4 shows the percentage of population in each per capita income range according to city size. There are five per capita income ranges, namely: poor (less

Table 3
Victimization per city size

% of population	City Size (thousands)			
	50 to 100	100 to 500	500 to 1,000	More than 1,000
Burglary/Larceny	3.416	5.558	7.932	7.945

than one minimum wage), lower middle class (between one and three minimum wages), middle class (between three and five minimum wages), upper middle class (between five and eight minimum wages), and rich (more than eight minimum wages). Notice that in larger cities, the share of wealthier people is larger. And from Table 3, there is an increasing relationship between city size and violence. So, the higher share of rich people in large cities might be behind the stylized fact that larger cities have higher crime rates. In the next section we will investigate this and other stylized facts in further detail.

Table 4
Percentage of population per income range and city size

Income	City Size (thousands)			
	50 to 100	100 to 500	500 to 1,000	More than 1,000
Poor	25.726	23.957	21.943	20.365
Lower middle class	59.095	57.355	51.197	47.907
Middle class	11.668	13.154	16.828	16.597
Upper middle class	1.623	3.288	5.822	7.863
Rich	1.886	2.244	4.208	7.266

4. Estimations and Results

Our purpose in this section is to assess the determinants of the likelihood of becoming a victim of burglary or larceny in São Paulo state, Brazil. The victimization outcome provided by the survey is of discrete nature, i.e., a person is or is not a victim. This suggests the use of the probit model technique. The baseline specification to be estimated using the survey sample weights is the following:

$$\begin{aligned} Y^*_{ihr} &= \beta_0 + X_i\beta_1 + Z_h\beta_2 + \gamma_r\beta_3 + \varepsilon_{ihr} \\ Y_{ihr} &= 1 \quad \text{if} \quad Y^*_{ihr} > 0, Y_{ihr} = 0 \quad \text{if} \quad Y^*_{ihr} \leq 0 \end{aligned} \tag{3}$$

where Y_{ihr} is a dummy variable indicating whether individual i who lives in domicile h in region r was a burglary/larceny victim ($Y^*_{ihr} > 0$) or not ($Y^*_{ihr} \leq 0$). X_i is a vector of the individual characteristics (income, age, race, education, etc), Z_h is a vector of the domicile characteristics (number of dwellers, marginal dwelling, etc). γ_r is a vector of São Paulo state Metropolitan Area indicators and ε_{ihr} is the unobserved error term.

The use of sample weights accounts for the different group proportions in the population. For example, there are more young victims because there are more youngsters in the population, but they would receive a smaller sample weight in order to account for oversampling.

A major difficulty in relating the theoretical model to the data is to have explanatory variables that capture just one of the factors influencing the victimization likelihood, namely exposure, proximity, attractiveness and self-protection. For example, income is without a doubt related to attractiveness, but also rich people would invest more in self-security and would go to different places, thus income also affects exposure and proximity. Age is one determinant of exposure since there are some places in which there are only young or old people. On the other hand, a younger person might be less attractive since income tends to increase with age. Given this problem, we opted to include, in both X_i and Z_h, variables that are related to at least one of these four factors, such as the ones discussed earlier in this paragraph.

An important feature of the law enforcement system in Brazil is that criminal laws are the same for all states, and the police and criminal justice system are established at the state-level. Since our data encompass only São Paulo state, we do not need to control for differences in police and judiciary systems.

The first set of regressions intends to estimate the impact of city size, metropolitan area, and city growth on the likelihood of becoming a victim of burglary/larceny. Table 5 shows the regression output, in which the omitted category for the city size indicators is cities with less than 100,000 people.

We found positive and statistically significant city size coefficients, which imply a positive correlation between city size and victimization, in line with the result found by GS and GP. However, contrary to GP findings, the relationship between city size and likelihood of being a victim is not increasing, i.e., the likelihood is larger for the 500,000 to 1,000,000 population range than for cities with more than one million inhabitants, although the likelihood for both ranges are larger in relation to smaller city size ranges.

Additionally, we found that living in Santos MA increases the victimization probability, but the same is not true for São Paulo MA. Both MA contain cities of several sizes, having at least one city in each range, so the MA indicator may not be capturing non-linearities of the city size effect. Contrary to GP findings of a positive and statistically significant city size growth coefficient, which was a cornerstone of their paper, we found that this coefficient was negative and statistically significant. If this coefficient captures the law enforcement congestion effect hypothesized by GP, it seems that either this was not the case for São Paulo state cities, or the congestion effect occurred several years before 1998, and now the coefficient captures the government response to the previous increase in crime.

According to GS, in the USA the likelihood of burglary/larceny victimization for an individual living in a city with more than one million inhabitants is 28%

larger than in a city in the 50,000 to 100,000 range.[9] For GP, this figure was 78%, and for us it is 82%. It seems that for Latin American countries the difference in violence between big and small cities is very large in comparison to U.S. estimates.

Table 5
Probit estimates of burglary/larceny victimization – City size and population

Variable	(1) Burglary Larceny	(2) Burglary Larceny	(3) Burglary Larceny
Population between 100,000 and 500,000	0.0304* (0.0065)	0.0268* (0.0064)	0.02540* (0.0064)
Population between 500,000 and 1,000,000	0.0659* (0.0115)	0.0642* (0.0117)	0.0546* (0.0110)
Population larger than 1,000,000	0.0554* (0.0073)	0.0515* (0.0089)	0.0403* (0.0091)
Santos MA		0.0371* (0.0067)	0.0342* (0.0067)
São Paulo MA		0.0062 (0.0054)	0.0093 (0.0055)
City growth			-0.2773* (0.1072)
Number of obs.	28583	28583	28583
Pseudo R^2	0.0083	0.0098	0.0105

Note: Robust standard errors reported in parentheses, marginal effects evaluated at the mean.
*means statistically significant at the 5% level.

A hypothesis upon which our theoretical model rests is that individuals take the city size as given. Nevertheless, if the individual's choices about the size of the city of residence take into account her chance of being victimized, our estimates of city size effect would not be consistent. Because if individuals whose characteristics imply a higher likelihood of victimization choose to live in larger cities, then one would find a positive relationship between city size and victimization that is not due to the size of the city per se. This possibility was first pointed out by GS, but they did not suggest any solution.

We could test for the endogeneity of location with our data using Evans et al. (1992) methodology. The first step consists in estimating a simultaneous equation model where the first equation is a probit specification of the victimization likelihood with the city size and the individual characteristics as explanatory variables. The other equation is a linear regression of city size on the likelihood of victimization of the individual and on the characteristics of the household head who, we believe, is in charge of choosing the city size. More information on this type of model and its estimation is available in Keshk (2003).

[9]Details about this calculation are described in GS paper.

$$
\begin{aligned}
Y_{ihr} &= \beta_0 + X_i\beta_1 + Z_h\beta_2 + \gamma_r\beta_3 + \text{city size } \beta_4 + \varepsilon_{ihr} \\
\text{city size} &= \lambda_0 + X_{ih}^H\lambda_1 + Z_h\lambda_2 + Y_{ihr}\lambda_3 + u_{ih}
\end{aligned} \tag{4}
$$

where X_{ih}^H contains the household head's characteristics.

Table 6 shows the output of the estimations using the city size variable expressed in population size (gpop98), the baseline single equation probit model (on the left) and the estimated coefficients of the first equation of the system (on the right). For the second step of the test, we conducted a Hausman specification test where the null hypothesis is that the single equation probit provides consistent and efficient estimates, whereas the simultaneous equation is consistent under the null and the alternative hypotheses, but is not efficient under the null. The test statistic is displayed below.

$$
\begin{aligned}
H &= \left(\hat{\beta}_{Simulteq} - \hat{\beta}_{Probit}\right)^2 / \left(Var\left(\hat{\beta}_{Simulteq}\right) - Var\left(\hat{\beta}_{Probit}\right)\right) \\
H &= \left(2.46 \times 10^{-8} - 1.34 \times 10^{-8}\right)^2 / \left[\left(7.64 \times 10^{-9}\right)^2 - \left(3.42 \times 10^{-9}\right)^2\right] \\
&= 2.688
\end{aligned}
$$

where β is the estimated coefficient for the city size variable. H is distributed as a chi-square with one degree of freedom. Its p-value is 0.1016, so we are not able to reject at the 5% level of confidence that the single equation probit generates consistent estimates.

As a robustness check, we conducted the same test, but now using the city size variable expressed in the log of population size.[10] The Hausman test statistic for this case is $H = 1.12$, with a p-value of 0.28, and thus the null hypothesis cannot be rejected again. Given these results, we decided to stick with the single equation probit models in the next estimations.

Then, we will investigated if individual characteristics play an important role per se and through city size in determining the victimization likelihood. To do that, we ran regressions adding controls for individual characteristics, as reported in Table 7.

The first regression included demographic variables such as age, gender, color, origin, employment, family characteristics and income ranges. In the second regression, we added variables related to education, in order to account for both different lifestyles related to education and to control for a possible understatement of income. And the third regression includes housing-related variables such as number of dwellers, public utilities and if the family lives in an apartment, in addition to all the previous variables.

[10] These estimates are available upon request.

Table 6
Baseline probit model and simultaneous equation model using city size

Model	Probit		Simultaneous equation	
Regressors	Coefficient	SE	Coefficient	SE
City population in 1998	1.35E-08**	3.35E-09	2.35E-08**	7.69E-09
Age 16 to 24 years	0.4355**	0.0570	0.3649**	0.0450
Age 25 to 34 years	0.4714**	0.0583	0.3815**	0.0463
Age 35 to 44 years	0.4434**	0.0604	0.3692**	0.0473
Age 45 to 59 years	0.3666**	0.0644	0.2819**	0.0501
Age over 60 years	0.2671**	0.0766	0.1221**	0.0597
Male	0.2455**	0.0308	0.2592**	0.0249
Black	-0.0500	0.0388	-0.0499	0.0319
Asian	0.0366	0.1405	-0.0977	0.1206
Foreigner	0.0341	0.1450	0.0690	0.1235
Migrant	-0.0347	0.0340	0.0114	0.0272
Occupied	0.0795	0.0507	0.0865**	0.0399
Couple	-0.3144**	0.0740	-0.2999**	0.0570
Couple with children	-0.1923**	0.0629	-0.1833**	0.0496
Single-parent family	-0.0040	0.0687	-0.0377	0.0540
Car	0.0722	0.0373	0.0526	0.0296
Health insurance	0.0160	0.0353	0.0050	0.0277
Lower middle class	0.0554	0.0445	0.0220	0.0354
Middle class	0.1532**	0.0606	0.1400**	0.0484
Upper middle class	0.3733**	0.0805	0.2746**	0.0669
Rich	0.2881**	0.0872	0.2809**	0.0737
Number of dwellers	-0.0397**	0.0117	-0.0334**	0.0091
Public lighting	-0.0086	0.0930	0.0754	0.0808
Not a visitor	0.0008	0.1206	-0.0966	0.0925
Marginal dwelling	-0.0915	0.0472	-0.0272	0.0371
Own home	-0.0353	0.0360	-0.0501	0.0290
Living in a seized area	-0.0065	0.0794	-0.0542	0.0709
Constant	-1.8630**	0.1780	-1.8261**	0.1446
Observations	27,959			27,959

Note: **means statistical significance at the 5% level.

City size is still an important determinant of victimization since in both sets of regressions not only did all coefficients keep the same signal and magnitude, but they also were statistically significant. Interestingly, at this time, the São Paulo MA variable was statistically significant and positive, and its marginal effect was larger than the one for Santos MA. The city growth variable remained negative but not statistically significant.

Table 7
Probit estimates of burglary/larceny victimization – lifestyle controls

Model	(1)		(2)		(3)	
Independent variables	Coefficient (dF/dx)	SE	Coefficient (dF/dx)	SE	Coefficient (dF/dx)	SE
Population between 100k – 500k	0.0222**	0.006	0.0214**	0.006	0.0226**	0.006
Population between 500k – 1M	0.0437**	0.010	0.0412**	0.010	0.0435**	0.010
Population larger than 1M	0.0285**	0.008	0.0265**	0.008	0.0286**	0.008
São Paulo MA	0.0363**	0.007	0.0332**	0.007	0.0381**	0.008
Santos MA	0.0130**	0.013	0.0134**	0.005	0.0139**	0.008
City growth	-0.1324	0.099	-0.1116	0.098	-0.0835	0.101
Age 16 to 24 years	0.0618**	0.009	0.0390**	0.009	0.0400**	0.009
Age 25 to 34 years	0.0719**	0.010	0.0509**	0.010	0.0481**	0.010
Age 35 to 44 years	0.0674**	0.011	0.0499**	0.010	0.0476**	0.010
Age 45 to 59 years	0.0541**	0.011	0.0475**	0.010	0.0444**	0.010
Age over 60 years	0.0394**	0.012	0.0417**	0.012	0.0366**	0.012
Male	0.0284**	0.004	0.0284**	0.004	0.0288**	0.004
Black/multiracial	-0.0091**	0.004	-0.0065	0.004	-0.0052	0.004
Asian	0.0019	0.016	0.0001	0.016	0.0007	0.016
Foreigner	0.0044	0.017	-0.0019	0.015	-0.0046	0.015
Migrant	-0.0039	0.004	-0.0002	0.004	0.0001	0.004
Employed or occupied	0.0083	0.005	0.0070	0.005	0.0082	0.005
Couple only	-0.0305**	0.005	-0.0304**	0.005	-0.0294**	0.006
Couple with children	-0.0352**	0.008	-0.0351**	0.008	-0.0233**	0.008
Single-parent family	-0.0083	0.007	-0.0092	0.007	-0.0025	0.008
Car	0.0096**	0.004	0.0066	0.004	0.0055	0.004
Lower middle class	0.0107**	0.005	0.0065	0.005	0.0021	0.005
Middle class	0.0250**	0.008	0.0132*	0.007	0.0063	0.007
Upper middle class	0.0642**	0.014	0.0442**	0.014	0.0341**	0.013
Rich	0.0495**	0.014	0.0284**	0.013	0.0188	0.013
Schooling 5 to 8 years			0.0255**	0.005	0.0241**	0.005
Schooling 9 to 12 years			0.0381**	0.007	0.0365**	0.007
Schooling more than 12 years			0.0422**	0.010	0.0407**	0.011
Marginal dwelling					-0.0097**	0.005
Number of dwellers					-0.0045**	0.001
Apartment					0.0003	0.006
Pseudo R-squared	0.0488		0.0540		0.0565	
Observations	28583		28583		28583	

Note: Marginal effects reported are evaluated at the mean of the regressors. The dF/dx for dummy variable is a discrete change from 0 to 1, Robust standard errors were used.
All models included a constant;
z and $P > |z|$ are the test of the underlying coefficient being equal to zero.
100k is 100,000 people and 1M is 1,000,000 people.
** means statistically significant at 5%.

The lifestyle-related variables also seem to have an important role. The age categorical variables, in which 0-15 year category was omitted, were positive and tended to be declining as age increased, which is also reported in GS and GP. Gender was also important, because men are more likely to be victims. Black and/or other racial background was significant in the first regression but not in the others, in which education and housing variables were added. The variables related to origin and employment coefficients were not statistically significant. The family status, in which single is the omitted category, had a significant marginal effect, being negative for married people (couples) having or not children, although for single-parent families the effect was not statistically significant.

In the first regression of Table 7 the income categorical variables (poor was the omitted category) were positive and statistically significant. Their marginal effects were increasing up to the upper middle-class range, being the effect for rich people smaller than the one for the upper middle class. Another variable included that could capture both income- and lifestyle-related factors was the family ownership of a car. This variable coefficient was positive and statistically significant only in the first regression.

In the second regression of Table 7 we added the education categorical variables (0-4 years of education was the omitted category), which presented statistically significant coefficients. However, this inclusion made the variables black, car, and lower middle-class income lose their statistical significance. The marginal effects of income variables changed in magnitude, but they kept the same pattern discussed before. The pseudo-R^2 increased by more than 10% from the first to the second column, and it indicates a better fit to the data.

Adding housing-related variables in the third regression did not change the statistical significance of the previous variables, except for the income variables, which declined more in absolute terms, and only the upper middle class remained significant. The apartment dummy variable was not statistically significant. But the number of persons living in the same home was statistically significant and had a negative sign, i.e., the larger the number of dwellers, the smaller the likelihood of becoming a victim. Maybe this variable could be capturing poverty, because not only do poor people tend to have more children, but also several generations of the same family live in the same house. The last dummy variable included was marginal dwelling, which is "one" if the street where the home is located is not paved, does not have public lighting, garbage collection, electricity, water or sewage service, and "zero", otherwise. Its coefficient was negative and statistically significant, which means that poor people living in marginal dwellings are less likely to be victims of burglary/larceny. The pseudo-R^2 presented a slight increase from the last regression, so the new variables did not improve significantly the model fit to the data, which can also indicate that the new variables are highly correlated to the old ones.

We could not reject the null hypothesis that the coefficients estimated for the upper middle class and the rich are statistically different for any specification of Table 7,[11] however, we believe that this fact deserves a more detailed examination.

The results from Table 7 indicate that the implied elasticity between city size and crime drops from 0.16 (in the last model of Table 5) to 0.041 (in the last model

[11] The p-value of this test for column (3) is 0.265.

of Table 6), a fall of 75%, in contrast to the 33% decline found by GS and 8% decrease found by GP. Thus, contrary to both previous papers, a large share of the city size effect is accounted for by the differences among the cities' household characteristics (see GP, p. 192).

The regression results pose some safety policy issues. The first is that since middle-class and rich people are more likely to be victims of burglary/larceny, they will need more police protection than poor people. In terms of public policy it would be interesting to know if the effect of income depends on the size of the city where the possible victim lives. Actually, in our theoretical model there is an interaction between income and city size through the investment in self-protection. Hence, we ran regressions using the interaction between income categories and city size. The results for burglary/larceny are reported in Table 8.

The first column of Table 8 contains the regression in which the "rich" dummy variable was replaced with its interaction with the city size categories. The second regression had the "upper middle class" dummy replaced with its interaction with the city size categories, and so on. Contrary to GP, the regression results for burglary/larceny show that people in different income ranges are differently affected according to city size. Interestingly, when contrasting Table 8 regressions with column 1 of Table 7, we can see that the pseudo-R^2 increased by at least 15%, which indicates that adding the interaction between city size and income to the regression significantly improved the fitting of the model.

From the "rich people" regression, we can see that rich people are more likely to become victims in cities with 100,000-500,000 and 500,000-1,000,000 inhabitants, and in the latter, the likelihood is more than five times higher than if they lived in cities with more than one million people, whose effect, although positive, was statistically significant only at the 10% level. On the other hand, upper middle-class people have more chance to become victims in cities with more than one million inhabitants, which was the only statistically significant city size range. Middle-class people are more likely to be victimized in cities with more than 500,000 inhabitants, but the largest effect is obtained for cities in the 500,000-1,000,000 range, whereas the effect for cities with less than 500,000 is positive. All of the interaction effects for the lower middle-class income range are not statistically significant at the 5% level as was also the case with the lower middle-class income dummy variable.

GP proposed the interesting hypothesis that big cities have a larger share of crime-prone individuals than smaller cities. We cannot reject this hypothesis because from Table 4 we can see that the larger the city, the larger the proportion of rich and upper middle-class people in the overall population. This lures crime-prone individuals into large cities, and the rate of crime against rich people should be higher than in smaller cities, or else equal.

An interesting fact uncovered by us is the decrease in the likelihood of rich people being victims of burglary/larceny in relation to the upper middle class.

Among the possible explanations, we can think of three reasons:

i) The rich are afforded more police protection,

ii) Rich people adopt a middle-class lifestyle to deceive offenders, and

iii) Rich people can afford to invest in self-security by hiring bodyguards, using armored cars, etc.

If the first reason is the case, rich people would also be protected by the police in smaller cities, because if rich people can influence the police in large cities, there is no reason for them not to do so in small cities. But our results rule out this reason. The second reason does not appear to be corroborated by our results either, because if the deception strategy were successful, the upper middle class could adopt it too.

We were not able to rule out the third reason based on our results, so we are led to believe that this decrease in rich people's victimization likelihood might be due at least in part to investment in self-protection. In such case, our theoretical model suggests a way to measure an upper bound of the effect on rich people's victimization likelihood by changing the investment in self-protection from a city with more than 1,000,000 inhabitants to a city with 500,000-1,000,000 inhabitants. This effect is the difference between the coefficient of the interaction dummy variable for a rich person living in a 500,000-1,000,000 city and the interaction for a rich person living in a city with more than 1,000,000 inhabitants:[12] $14.39\% - 3.48\% = 10.91\%$.

This result should be taken with a grain of salt because not only it is an upper bound, but also there are other factors related to city size not contemplated by the theoretical model that can affect the victimization pattern among cities. Furthermore, since our dataset does not contain information about private investment in self-protection, we are not able to test it. And in fact, the only victimization survey we know of that contains some information on private investment in self-protection was conducted in 2002 and 2006 by CRISP-UFMG for the city of Belo Horizonte.[13] Nevertheless, this survey does not provide enough detailed information to uncover temporal (investment was done before or after victimization) and spatial relationships (e.g., if neighbors are installing a fence, I should also install one because my house is now relatively less protected than my neighbor's) between crime and investment in private security. After all, the effect of private investment in security is still an open empirical question.

[12]The difference between the estimated coefficients is statistically different from zero at the 5% level.

[13]There is a victimization survey conducted in 2002 and 2006 by CRISP-UFMG, see www.crisp.ufmg.br

Table 8
Probit estimates of burglary/larceny victimization – City and income interactions

Income range	Rich		Upper middle class		Middle class		Lower middle class	
Independent Variables	Coefficient (dF/dx)	SE	Coefficient (dF/dx)	SE	Coefficient (dF/dx)	SE	Coefficient (dF/dx)	SE
Population between 100k – 500k	0.0201**	0.006	0.0219**	0.006	0.0185**	0.006	0.0251**	0.007
Population between 500k – 1M	0.0355**	0.009	0.0424**	0.010	0.0360**	0.010	0.0411**	0.012
Population larger than 1M	0.0261**	0.008	0.0241**	0.008	0.0239**	0.008	0.0220**	0.008
São Paulo MA	0.0328**	0.006	0.0338**	0.006	0.0323**	0.006	0.0326**	0.006
Santos MA	0.0136**	0.005	0.013**	0.005	0.0129**	0.005	0.0135**	0.005
City growth	-0.0947	0.100	-0.130	0.101	-0.0679	0.103	-0.1724	0.129
Age 16 to 24 years	0.0390**	0.009	0.0389**	0.009	0.0389**	0.009	0.0391**	0.009
Age 25 to 34 years	0.0505**	0.009	0.0510**	0.010	0.0507**	0.010	0.0510**	0.010
Age 35 to 44 years	0.0500**	0.010	0.0494**	0.010	0.0497**	0.010	0.0501**	0.010
Age 45 to 59 years	0.0475**	0.010	0.0474**	0.010	0.0473**	0.010	0.0477**	0.010
Age over 60 years	0.0421**	0.012	0.0417**	0.012	0.0414**	0.012	0.0418**	0.010
Schooling 5 to 8 years	0.0257**	0.005	0.0256**	0.005	0.0257**	0.005	0.0255**	0.005
Schooling 9 to 12 years	0.0383**	0.007	0.0382**	0.007	0.0385**	0.007	0.0382**	0.007
Schooling more than 12 years	0.0427**	0.010	0.0428**	0.010	0.0424**	0.010	0.0423**	0.010
Lower middle class	0.0062	0.004	0.0064	0.005	0.0062	0.004	-	-
Middle class	0.0128*	0.007	0.0132*	0.007	-	-	0.0119*	0.007
Upper middle class	0.0435**	0.013	-	-	0.0445**	0.013	0.0435**	0.013
Rich	-	-	0.0288**	0.013	0.0287**	0.013	0.0281*	0.013
income * pop between 100k-500k	0.0526**	0.030	-0.0077	0.017	0.0244*	0.014	-0.0060	0.008
income*pop between 500k-1M	0.1439**	0.059	-0.0014	0.023	0.0332*	0.019	-0.0001	0.160
income * population larger than 1M	0.0348*	0.021	0.0491**	0.020	0.0247*	0.012	0.0086	0.007
income * City growth	-0.6016	0.431	0.4862	0.381	-0.3312	0.232	0.1111	0.160
Pseudo R-squared	0.0549		0.0549		0.0549		0.0542	
Observations	28583		28583		28583		28583	

Note: Marginal effects reported are evaluated at the mean of the regressors.
The dF/dx for dummy variable is a discrete change from 0 to 1;
z and $P > |z|$ are the test statistic of the underlying coefficient being equal to zero.
100k is 100,000 people and 1M is 1,000,000 people.
* and ** mean statistically significant at 10 and 5%, respectively.
Other controls used: Male, Black/other, Asian, Foreigner, Migrant,
Employed or Occupied, Couple only, Couple with children, Single-parent family, Car.

If, at least in part, this reduction in victimization likelihood is due to the investment in self-protection, one might ask why other people do not invest in it. The answer should be that self-protection investment is so expensive that only rich people that live in cities with more than 500,000 inhabitants may find it worthwhile, i.e. the investment cost is smaller than the expected loss due to burglary/larceny.

In our previous analysis, two important topics were left out. The first limitation is about the nature of the dependent variable that is binary (yes/no) instead of a count variable, i.e. number of burglary/larceny suffered. Thus, we should expect biased estimated coefficients; however, the direction of the bias is not clear.

The second topic is that our cross-sectional data do not capture changes in crime over time. Suppose that cities with 500,000 and 1,000,000 inhabitants were safer before 1998. Then, given the investment in self-protection in large cities, criminals might get a smaller return for crime and would migrate to smaller cities where people are less protected. Our data might be capturing this exact moment. Later on, the increase in the expected loss of potential targets in medium-sized cities would induce them to invest in self-protection; therefore the crime rate would eventually decline in these regions, and the increasing relation between crime and city size would be re-established.

5. Conclusions

In this paper, we present a theoretical model and a detailed empirical assessment of the determinants of burglary and larceny victimization in São Paulo state. By taking into account the potential victims' investment in self-security and criminals with heterogeneous ability, our theoretical model can generate a non-increasing relationship between the income of an individual and her victimization likelihood.

Our empirical results do not reject our theoretical model hypothesis that individuals do not take into account their larceny/burglary victimization likelihood when choosing a city to live in. We find that city size is an important factor in the likelihood of being a victim of burglary/larceny; however, contrary to the previous literature, most of this effect is accounted for by the potential victim's characteristics such as income, education, etc. Additionally, recent city growth seems to have no role at all in explaining the likelihood of being a victim. If congestion effects on law enforcement happened, it occurred several years before our data were collected.

City size and crime are not increasing even when conditioning them on the income of the potential victim. In addition, contrary to previous findings, the data provided evidence that the likelihood of being a victim is not monotonically increasing in the victim's income, as predicted by our theoretical model. We suspect that one of the reasons behind this finding could be investment in self-protection. Therefore investments in self-protection might have an important role that has been ignored so far, and this issue certainly deserves deeper investigation.

References

Beato, C., Peixoto, B., & Andrade, M. (2004). Crime, oportunidade, vitimização. *Revista Brasileira de Ciências Sociais*, 19(55):73–89.

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

Bursik, R. & Grasmick, H. (1993). *Neighborhoods and Crime*. Lexington Books, San Francisco, CA.

Catão, Y. (2000). Pesquisas de vitimização. Anais do Seminário Fórum de Debates Criminalidade, Violência e Segurança Pública no Brasil: Uma Discussão sobre Bases de Dados e Questões Metodológicas, Rio de Janeiro. Available: www.candidomendes.br/cesec/atividades/seminarios/f_debates.asp.

Clarke, R. & Felson, M. (1993). *Routine Activity and Rational Choice*. Transaction Publishers, New Brunswick, London.

Cohen, L. & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 44:588–608.

Cohen, L. E., Kluegel, J. R., & Land, K. C. (1981). Social inequality and predatory criminal victimization: An exposition and test of a formal theory. *American Sociological Review*, 46:505–524.

CRISP (2002). Pesquisa de vitimização em Belo Horizonte. Mimeo.

Evans, B., Oates, W., & Schwab, R. (1992). Measuring peer group effects: A study of teenager behavior. *Journal of Political Economy*, 100(5):966–991.

Fajnzylber, P. & Araújo Jr., A. (2001). Violência e criminalidade. In Menezes, N. & Lisboa, M., editors, *Microeconomia Aplicada no Brasil*. Fundação Getulio Vargas, Rio de Janeiro.

Fajnzylber, P., Lederman, D., & Loayza, N. (2001). *Crimen y Violencia en América Latina*. Alfaomega and World Bank, Bogotá and Washington.

Felson, M. & Cohen, L. (1980). Human ecology and crime: A routine activity approach. *Human Ecology*, 8:389–406.

Garafalo, J. (1987). Reassessing the lifestyle model of criminal victimization. In Gottfredson, M. R. & Hirschi, T., editors, *Positive Criminology*, pages 23–42. Sage, Beverly Hills, CA.

Gaviria, A. & Pagés, C. (2002). Patterns of crime victimization in Latin America cities. *Journal of Development Economics*, 67:181–203.

Glaeser, E. L. & Sacerdote, B. (1996). Why is there more crime in cities? NBER Working Paper 5430.

Gottfredson, M. R. & Hindelang, M. J. (1981). Sociological aspects of criminal victimization. *Annual Review of Sociology*, 7:107–128.

Hindelang, M., Gottfredson, M. R., & Garafolo, J. (1978). *Victims of Personal Crime: An Empirical Foundation for a Theory of Personal Victimization.* Ballinger Publishing Company, Cambridge, MA.

Jensen, G. & Brownfield, D. (1986). Gender, lifestyles and victimization: Beyond routine activity theory. *Violence and Victims*, 1:85–89.

Keshk, O. (2003). CDSIMEQ: A program to implement two-stage probit least squares. *The Stata Journal*, 3(2):1–11.

Khan, T. (1999). Os custos da violência: Quanto se gasta ou deixa de ganhar por causa do crime no Estado de São Paulo. *São Paulo em Perspectiva*, 13(4):42–48.

Kilsztjn, S., Rossbach, A., Carmo, M. S. N., Sugahara, G. T. L., & Souza, L. B. (2003). Vítimas fatais da violência e mercado de drogas na região metropolitana de São Paulo. *Revista Brasileira de Estatística Populacional*, 20(2):259–279.

Newman, O. (1972). *Defensible Space.* MacMillan, New York.

Peixoto, B., Andrade, M., & Moro, S. (2007). Violência urbana: Uma análise comparativa da vitimização em São Paulo, Rio de Janeiro, Recife e Vitória. Texto para Discussão 323.

Piquet, L. & Fajnzylber, P. (2001). A criminalidade nas regiões metropolitanas do Rio de Janeiro e São Paulo: Determinantes da vitimização e políticas públicas. In Fajnzylber, P., Lederman, D., & Loayza, N., editors, *Crimen y Violencia en America Latina.* Banco Mundial and Alfaomega.

Reppeto, T. A. (1974). *Residential Crime.* Ballinger, Cambridge, MA.

SEADE (1999). Segurança. Cadernos do Fórum São Paulo – Século XXI – Caderno 7. São Paulo GESP/SEADE. Disponível em: http://www.al.sp.gov.br/web/forum/CADERNOS/index.htm.

SEADE (2004). São Paulo: Fundação sistema estadual de análise de dados. Disponível em www.seade.sp.gov.br.

Soares, R. R. (2004). Development, crime and punishment: Accounting for the international differences in crime rates. *Journal of Development Economics*, 73:155–184.