Income and Repeat Criminal Victimization in Brazil

Marcelo Justus[1,*], Tulio Kahn[2], Henrique Kawamura[3]

Abstract

We analysed the effect of income on repeat criminal victimization in Brazil using data from the 2009 National Household Sample Survey and its special supplement on public safety. Two count-data models were estimated for four types of crime: theft, robbery, attempted theft/robbery, and assault. We found a nonlinear positive effect of income on repeat victimization for the three types of property crimes, and a nonlinear negative effect of income on assault.
Keywords: repeat victimization, crime, violence, income

Resumo

Analisamos o efeito da renda na vitimização criminal repetida no Brasil utilizando dados da PNAD 2009 e do suplemento especial sobre segurança pública. Foram estimados dois modelos de dados de contagem para quatro tipos de crime: furto, roubo, tentativa de furto/roubo e agressão física. Concluiu-se que a renda causa um efeito não-linear positivo na vitimização repetida para os três tipos de crimes contra a propriedade, e um efeito não-linear negativo na agressão física.
Palavras-chave: vitimização repetida, crime, violência, renda

Área ANPEC: Economia Social e Demografia
JEL Classification: K42, C25

1. Introduction

In the criminal universe, phenomena rarely follow a normal distribution. It is rather a universe governed by "concentrations." A small part of the territory tends to accumulate a large proportion of crime (hot spots). A small number of criminals tend to commit a disproportionate amount of crime (predators). Victimization also follows this trend, as a small group of victims is usually the preferred target of a disproportionate amount of offenses.

Situational criminology suggests that these concentrations are explained by a combination of excessive risk factors and the absence of protective factors. An unmonitored area with intense circulation of people and goods, low visibility, and signs of disorder becomes more attractive for the commission of crimes. Low-weight infants, living in homes with unmarried teenage mothers, raised by lone parents, or who leave school early are more likely to become criminals. Unmarried young people, who often tend to leave their home unguarded, consume alcohol, and be careless with their fancy phones are preferred prey for criminals.

From the standpoint of public crime prevention policies, these concentrations are advantageous, as they make it possible for resources to be allocated to areas and populations at risk through focused interventions, reducing risk factors and increasing protective factors. Identified hot spots can

*Corresponding author.
Email addresses: marcejustus@eco.unicamp.br (Marcelo Justus), concrim@uol.com.br (Tulio Kahn), henriquemura@gmail.com (Henrique Kawamura)
[1]Institute of Economics, University of Campinas.
[2]PhD in Political Science.
[3]Latin American Institute of Economics, Society and Politics, Federal University of Latin American Integration.

be more protected by police or cameras, better lighting, better cleaning services. Tertiary prevention programs – designed for people already involved with the criminal justice system – can focus on strengthening educational and therapeutic treatments for young offenders. Police departments can develop courses, booklets and guidelines for preventive measures to be adopted by owners of establishments that are attacked often. The problem is that it is difficult, or even impossible, to change many risk factors. You can improve the surveillance of an area, but a central shopping promenade might be always used for the same purposes and activities, implying risk. You can improve the employability of young offenders, but you cannot modify their age, gender, IQ, or their involvement in crime in the past. Victims can change risk behaviors and install security equipment, but there are intrinsic characteristics of location, lifestyle, and architectural design, among others, which cannot be modified. That is why increasing protective factors can shift some crimes to other areas or victims, but some of them will inevitably remain concentrated in the same locations and targets.

Many previous studies analyzed why some locations are more attractive to criminals than others and why other ones concentrate certain types of crimes, while other studies investigated risk factors associated with criminal trajectories of repeat offenders. Very little has been written, however, about the phenomenon of repeat victimization (more than one of the same type of crime). Which variables could help us understand why, despite being a relatively rare phenomenon, victimization mainly affects a small percentage of victims?

Repeat victimization has some known characteristics. Most people are not victimized at all, but those who are present a high risk of being victimized again. Thus, prior victimization is one of the best indicators of future victimization. Moreover, recurrence can be rapid. In repeat victimization, the same type of criminal incident is experienced by the same victim or target within a specific period of time, as within a year, for example. Thus, repeat victimization refers to the total amount of offenses experienced by a victim or target, including initial and subsequent offenses.

Previous evidence of the causes of repeat victimization in Brazil was uniquely found by Carvalho and Lavor (2008) using data from a national survey carried out in 1988. The focus of their study was particularly on the effect of income inequality on property crime (composed of theft and robbery). Our study seeks empirical advances in the modelling of causes of repeat victimization. In particular, the main objetive of this paper is that of analyzing the effect of income on repeat criminal victimization by types of crime from an economic perspective.

Victimization is a complex process and, consequently, one that is difficult to be modeled empirically. There is a no single well-structured theory to guide empirical analyses in this field. Studies have usually been based on two approaches that consider victims as objects of study, highlighting the importance of their "lifestyle" and creation of "opportunities" for criminals to carry out their crimes. Empirical analyses have been mainly based on the theoretical framework proposed by Cohen et al. (1981). Using data from some previous studies, these authors expanded and formalized a sociological theory (which they refer to as the "opportunity model of predatory victimization") to explain victimization risk. According to this approach, there are five factors strongly related to risk: exposure, proximity, guardianship, target attractiveness, and definitional proprieties of specific crimes.

Some factors with a bearing on repeat victimization can have a different effect according to the type of crime in question, especially if the nature of the crime is considered, i.e. property crimes or crimes against a person. Income, in particular, is widely debated in the literature. We are accustomed to associating crime with poverty, which is true in connection with homicides and other violent crimes against a person. Thus, a negative relationship between income and victimization is plausible. In the case of property crimes, however, its effect is ambiguous. On the one hand, higher income reduces the propensity to engage in crime, but on the other it produces more attractive targets, as property crimes are primarily crimes of opportunity. The higher the income, the more goods a victim has, the greater the criminal opportunities. But this relationship is not necessarily linear: low-income individuals or places are less attractive, but after a certain threshold a higher income tends to increase defences against crime through strategies and equipment designed to "block opportunities," such as video cameras, alarms, and security devices. In short, wealthier individuals are on the one hand more economically attractive to criminals, but on the other they have stronger reasons and more money to spend on their own security, especially after their first victimization. Therefore, the effect of income on repeat victimization is ambiguous, but its net effect can be observed empirically.

We believe that the sociological approach cited above is also helpful as a framework to understand the process of repeat victimization, i.e. why some people are victims of the same type of crime twice or more times. Our analisys is also based on a simple victimization model proposed by Gaviria and Pagés (2002), where an individual's wealth is the focus. In particular, the hypotheses tested here are: 1) income causes a positive, albeit nonlinear, effect on repeat victimization; 2) income causes a nonlinear negative effect on repeat victimization. This paper intends to test these hypotheses using the most recent victimization data from a nationwide sample survey carried out in Brazil. Estimations for four types of crimes (theft, robbery, attempted theft/robbery, assault)[4] were performed separately. The design of complex surveys was taken into account, since ignoring the sample design tends to underestimate the actual variance. In summary, these are the main improvements made here in relation to the previous empirical study.

This paper is structured as follows: Section 2 presents a brief description of a useful theoretical framework for discussing the effect of income on repeat victimization; Section 3 provides details about the empirical modelling; results are discussed in Section 4; and Section 5 concludes the study.

2. A Simple Victimization Model

Gaviria and Pagés (2002) proposed a simple victimization model that, together with the approach proposed by Cohen et al. (1981), is very useful for the empirical modelling that we'll do in the next section and for understanding the results presented in Section 4.

Justus and Kassouf (2013) summarized the model's framework. There are two actors (citizens and criminals) and two stages. In the first stage, citizens (who are only different from one another according to their wealth level) decide how much they will spend on private protection. In the second stage, citizens are matched with criminals who in turn decide whether or not to commit a crime upon observing the wealth (w) of their prospective victims and their investments on private protection (e). Assuming that criminals make their decisions on the mere basis of pecuniary factors, we can say that they weigh two factors: if they are successful in committing the crime, they will be rewarded with a portion of the victim's wealth, given by α times w ($\alpha \leq 1$); and if they fail – the probability of which is p – they pay a penalty equivalent to F. According to Justus and Kassouf (2013) the authors haven't point this out, but one must consider that failure means that the criminal was accused, arrested, convicted, and punished as provided for in the law.

Three additional assumptions are made: the probability of being caught is assumed to increase monotonically with spending on private protection (i.e. $p = \mathbf{p}(e)$, where $\mathbf{p}' > 0$); victims and criminals are considered as risk-neutral; criminals are assumed to have complete information in that they observe their victim's wealth and are able to correctly infer their risks of being caught.

In this context, a criminal will attempt to victimize citizen i who possesses a wealth of w_i and spent e_i on private protection as long as the following inequality holds

$$(1 - \mathbf{p}[e_i])\alpha w_i - \mathbf{p}[e_i]F > 0 \tag{1}$$

Since all citizens are potential victims for criminals, a given citizen i can avoid becoming a victim if he or she spends at least h_i on private protection, where h_i indicates the spending on protection that would make a criminal indifferent between attempting to steal from i because the risk involved would be too high. In sum,

[4]Assault is an aggraveted physical agression against a person and the three other types are property crimes.

$$h_i = \mathbf{p}^{(-1)} \left[\frac{\alpha w_i}{\alpha w_i + F} \right] \tag{2}$$

where $\mathbf{p}^{(-1)}$ is the inverse of function $\mathbf{p}$ that links private spending on protection to the probability of a criminal being punished.

Equation(2) gives, for each wealth level, the minimal spending on private protection required to prevent crime by deterring criminals. Therefore, citizens must decide whether they will spend h_i on their own protection or will not invest on its at all. They will spend hi only if it does not exceed the prospective losses of being victimized. That is, if

$$h_i \leq \alpha w_i. \tag{3}$$

Wealthier persons would need, *ceteris paribus*, greater spending on private protection to avoid victimization. This is the conclusion reached based on the first derivative of Eq.(2) with respect to w,

$$\frac{\mathrm{d}h_i}{\mathrm{d}w_i} = \frac{\alpha F}{(F + \alpha w_i)^2 \mathbf{p}'[h]} > 0 \tag{4}$$

But are wealthier persons willing to spend more on protection to avoid being victimized? Or will they instead prefer to bear some risk? The answer depends on the second derivative of h with respect to w,

$$\frac{\mathrm{d}^2 h_i}{\mathrm{d}w_i^2} = -\frac{\alpha^2 F(2(F + \alpha w_i)\mathbf{p}'[h_i]^2 + F\mathbf{p}''[h_i])}{(F + \alpha w_i)^4 \mathbf{p}'[h_i]^3} \tag{5}$$

Equation (5) will be negative, unless the second derivative of $\mathbf{p}$ is both negative and large in absolute value. So the wealthy will routinely invest in private security to avoid being victimized unless $\mathbf{p}$ exhibits sharp diminishing returns to scale.

If the marginal returns of an extra amount spent on private protection against crime are very low, the wealthy will find it too expensive to reach the necessary level of protection to avoid being victimized and will rationally decide to bear some measure of risk. Otherwise, they will spend the portion of their wealth deemed necessary to avoid being victimized.

In summary, according to the approach adopted by Gaviria and Pagés (2002), the wealth of individuals determines both their economic attractiveness as victims and their capacity to protect themselves from criminals by paying for their protection.

Justus and Kassouf (2013) argue that according to the findings of Becker (1968), Ehrlich (1973), Cohen et al. (1981) and Gaviria and Pagés (2002) it is to be expected that, given the opportunity cost of crime, the likelihood of failure determined by government spending on public safety, the penalties provided for in the law, and the costs involved in planning and committing a crime, criminals will pick their victims based on their evaluation of those who are more economically attractive for the criminal act. In this subjective evaluation, criminals take into account both the wealth of potential victims and the likelihood of failure determined by how much they spend on their own protection. By doing this, criminals optimize the expected return on crime. Therefore, the behavior of potential victims has a direct bearing on the optimization process that is implicit in the rational choice of a criminal. Thus, if the principle of economic rationality on the part of criminals holds, the risk of victimization increases with wealth. However, as pointed out by Gaviria and Pagés (2002), wealthier individuals have stronger reasons and more money to spend on their own security to protect themselves from criminals. On the other hand, poorer individuals lack the financial means to pay for their protection to avoid being victimized, but are less economically attractive potential victims to criminals than wealthier individuals.

3. Empirical Modelling

The first difficulty in investigating the causes of crime is that reliable information is hard to come by (or virtually non-existent). Existing official data, especially that available in police records, consists only in underestimated figures for actually committed crimes. The number of criminal occurrences is underestimated because many of them are not actually registered in a formal police report. A police report involves assessments and decisions of various individuals involved in an event seen as a "police matter."

Actually, as suggested by victimization surveys conducted in several countries, there is clear evidence that the actual crime rate is significantly higher than that reported based on official data. A victimization survey is based on a random sample of a given population, which is asked about instances of certain types of crimes in a given period of time. Besides allowing for better measurement of the actual crime rate, among other advantages, these surveys make it possible for one to know the characteristics of the victims and provide important inputs for empirical studies on the causes of criminal victimization.

In this study, we used cross-sectional data from Special Supplements on Food Security, Victimization and Justice included in the National Household Sample Survey of 2009 (2009 PNAD in Brazilian acronym) carried out by the Brazilian Institute for Geography and Statistics (IBGE in Brazilian acronym). Our empirical models were estimated taking into account the design of complex surveys, since ignoring such sample design tends to lead to an underestimation of the actual variance.[5]

It must be said that our data set offers at least three advantages as compared to official crime figures: 1) its coverage is nation-wide; 2) the response variable (i.e., crime) is free from bias caused by measurement errors resulting from under-reporting, and 3) it allows for the effects of household income, education, and other factors on the number of crimes to be identified based on non-victimized individuals.

The response variable of the empirical models is a count-data: *the amount of times an individual was victimized during one year*. Models for four types of crimes were performed separately: robbery, theft, attempted theft/robbery, and assault. Table 1 shows the distribution of victimization count-data by type of crime.

Table 1: Frequency distributions of the count of victimizations by type of crime (%)

Count	Theft	Robbery	Attempted theft/robb.	Assault
0	96.09	96.37	94.67	98.47
1	3.02	2.86	3.95	1.13
2	0.57	0.53	0.91	0.20
3	0.19	0.16	0.30	0.09
4	0.06	0.04	0.08	0.03
5	0.04	0.03	0.05	0.04
6 or more	.04	0.02	0.05	0.05

$n = 318,774$.

The next step is also very difficult, as it is not trivial to specify a model without a well-structured crime causation theory. Fortunately, we can use the work of Cohen et al. (1981) and Gaviria and Pagés (2002) as a starting point.

We emphasize that for each explanatory variable selected there are several approaches that try and explain why and how risk of victimization are affected. For instance, men are more victimized because they tend to adopt risky behaviors due to the cultural roles assigned to them or to their higher testosterone levels. Whites and nonwhites tend to live in different areas of cities with different

[5] For details see Skinner et al. (1998).

levels of crime. Young people have a more active lifestyle that exposes them to greater risks, and married people tend to spend more time in their homes. Denser urban areas facilitate anonymity and make crime less detectable. People who study or work spend more time on the street, increasing their chances of victimization, and higher-income individuals carry more attractive goods than others.

In this study, income is the predictor of our greatest interest. Thus, we take into account the logarithm for monthly household income – `lln(inc)` – and also the number of individuals within a family (`famsize`) for property crimes (theft, robbery, attempted theft/robbery). For assault victimization, we decided to use the logarithm for monthly per capita household income – `ln(incp)`. Regarding types of crime, another difference is that for property crimes hours of work outside one's home in a week are used, while for assault only a dummy variable is used that assumes 1 if the individual is employed and 0 if otherwise (`works`).

Besides the 27 dummy variables used to take into account possible regional differences among Brazilian states, the other dependent variables that are common to the four models are: age in years (`age`) and the square of this variable (`agesq`); a dummy variable to distinguish gender, which assumes value 1 for males and 0 for females (`man`); a dummy variable to distinguish color or race (`white`), which assumes value 1 for white or Asian people and 0 for black, mulatto or indigenous people; a dummy variable for location of residence, which is 1 for urban areas and 0 otherwise (`urban`); a dummy for type of residence, which is 1 for house and 0 for apartament (`house`); a dummy variable for marital status, which is 1 for married and 0 for single (`married`); a dummy variable for student status, which is 1 for students and 0 otherwise (`student`); years of schooling (`school`); a dummy variable which is 1 if the person owns a car or motorcycle and 0 otherwise (`ownvehi`). Table A.1 describes all these variables precisely.

Because the dependent variable is discrete, its distribuition places probability mass at nonnegative integer values only. Fully parametric formulations of count models accommodate this property distribution. Cameron and Trivedi (2009) and Kleiber and Zeileis (2008) are convenient references for the methodological background for count-data model. A mix of both references was used in the next paragraphs.

The modeling exercise began with the standard linear regression model, which is estimated by OLS,

$$y_i = x_i'\beta + \varepsilon_i, \quad i = 1, ..., n, \tag{6}$$

where y_i is the number of times an individual i was victimized in one year, x_i is the column vector of covariates for observation i, which is described later, β is a $k \times 1$ vector of regression coefficients and ε_i is the error term.

There are three aspects in the linear regression model for a conditionally normally distributed response y: 1) linear predictor $\eta_i = x_i'\beta$, through which$\mu_i = E(y_i|x_i)$ depends on the $k \times 1$ vectors x_i of observations and β of parameters; 2) the distribution of the dependent variable $y_i|x_i$ is $N(\mu_i, \sigma^2)$; and 3) the expected response is equal to linear predictor, $\mu_i = \eta_i$.

The class of generalized linear models (GLMs) extends 1) and 3) to more general families of distributions for y and to more general relations between $E(y_i|x_i)$ and the linear predictor than the identity. Specifically, $(y_i|x_i)$ may now follow a density or probability mass function of the type

$$f(y; \theta, \phi) = \exp\left\{\frac{y\theta - b\theta}{\phi} + c(y; \phi)\right\} \tag{7}$$

where θ, referred to as the canonical parameter, depends on the linear predictor, and the additional parameter ϕ, referred to as the dispersion parameter, is often known. In addition, the linear predictor and the expectation of y are now related by a monotonic transformation (referred to as the link function of the GLM),

$$g(\mu_i) = \eta_i.$$

For fixed θ, (7) describes a linear exponential family. Then, the distribution of the dependent variable $y_i|x_i$ is a linear exponential family, a class that includes the Poisson and binomial distribution.

Thus, in addition to other possibilities, the family of GLMs extends the applicability of linear-model ideas to data where responses are binary or counts.

We start with a standard model for count data, which is a Poisson regression. As observed above, it is a generalized linear model. Using the canonical link for the Poisson family (the log link), the model is

$$E(y_i|x_i) = \mu_i = \exp(x_i'\beta) \tag{8}$$

In a Poisson distribution, the variance equals the mean (equidispersion). It is necessary to check this built-in feature. Poisson regressions are often plagued by overdispersion, which means that the variance is greater than the linear predictor permits.

Overdispersion can be tested for by considering the alternative hypothesis

$$\mathrm{Var}(y_i|x_i) = \mu_i + \alpha \cdot h(\mu_i), \tag{9}$$

where h is a positive function of μ_i. Overdispersion corresponds to $\alpha > 0$ and underdispersion to $\alpha < 0$.

A common specification of the transformation function h is $h(\mu) = \mu^2$ was used. Therefore, the formal test of the null hypothesis of equidispersion, $\mathrm{Var}(y_i|x_i) = E(y_i|x_i)$ for all i, against the alternative of overdispersion, is based on the equation

$$\mathrm{Var}(y_i|x_i) = E(y_i|x_i) + \alpha^2 E(y_i|x_i). \tag{10}$$

$H_0 : \alpha = 0$ was tested against$H_1 : \alpha > 0$ implemented by an auxiliary OLS regression of the generated dependent variable, $\{(y_i - \hat{\mu})^2 - y_i\}/\hat{\mu}$ on $\hat{\mu}$, without an intercept term, and performing a t test for whether the coefficient of $\hat{\mu}$ is zero, which is asymptotically standard normal under the null hypothesis.

The t statistics are 11.98 for theft, 8.72 for robbery, 10.29 for attempted theft/robbery, and 14.37 for assault, with one-sided p-value of 0.000 for all models. The results suggest that the Poisson model for the victimization data is not well specified, as there appears to be a substantial amount of overdispersion.

A possible solution is to consider a more flexible distribution that does not impose equality of mean and variance. The most widely used distribution in this context is the negative binomial. It may be considered a mixture distribution arising from a Poisson distribution with random scale, the latter following a gamma distribution. Its probability mass function is

$$f(y;\theta,\phi) = \frac{\mathit{\Gamma}(\theta + y)}{\Gamma(\theta)y!)} + \frac{\mu^y \theta^\theta}{(\mu+\theta)^{y+\theta}}, \quad y = 1, 2, \cdots, \mu > 0, \theta > 0. \tag{11}$$

It must be said that the variance of the negative binomial distribution is given by

$$\mathrm{Var}(y;\theta,\phi) = \mu + \frac{1}{\theta}\mu^2, \tag{12}$$

which is of the form (3) with $h(\mu) = \mu^2$ and $\alpha = 1/\theta$.

For estimating negative binomial model (NB) with known θ, the shape parameter of the fitted negative binomial distribution for the four models (i.e., types of crimes),$\hat{\theta}$, suggested that there is a considerable amount of overdispersion, corroborating with the results of the test for overdispersion.

A problem often faced with count data regressions is that the number of zeros is often much larger than a Poisson or negative binomial regression permits. In fact, Table 1 indicates that our data contains a large number of zeros. For instance, at least 94.7% of all people were not victimized. Thus, models for zero-inflated data are better.

The zero-inflated model was originally proposed to handle excess zeros relative to the Poisson model. It supplements a count density,$f_2(\cdot)$, with a binary process with a density of $f_1(\cdot)$. If the binary process takes on a value of 0 with a probability of $f_1(0)$, then $y = 0$. If the binary process takes on a value of 1 with a probability of $f_1(1)$, then y takes on the count values $0, 1, 2, \cdots$ from the count density $f_2(\cdot)$. This lets zero counts occur as a realization of the binary process and as a realization of the count process when the binary randon variable takes on a value of 1.

Suppressing regressors for simplicity, the zero-inflated model has a density of

$$f(y) = \begin{cases} f_1(0) + \{1 - f_1(0)\} f_2(0) & \text{if } y = 0 \\ \{1 - f_1(0)\} f_2(0) & \text{if} \, y \geq 0 \end{cases} \tag{13}$$

In this study, the $f_1(0)$ was parameterized through a Probit model and a negative binomial distribution for the count component was used. The zero-inflated negative binomial (ZINB) model is a mixture specification with a negative binomial count component and an additional point mass at zero.

We estimated a regression of victimization on all further variables for the count part and modelled the inflation component as a function of `man` and `white`. Apart from these variables, the previous study performed by Carvalho and Lavor (2008) takes into account a dummy variable for owning a TV set. Here, this control was not used because today almost all people have a TV set at their home. Moreover, the dummy variables for the Brazilian states were used as variables in the inflation component.

The LR test of Vuong to discriminate between the NB and ZINB models (see Cameron and Trivedi, 2009, 586-590) was applied. The test statistic is standard normally distributed, with large positive values favoring the ZINB model and large negative values favoring the NB model. In this study, the test statistic is a large positive value for the four models (i.e. types of crime) with a one-sided very small p-value. So only the results of ZINB models will be discussed next.

4. Results

Our model includes several variables that correlate significantly with repeat victimization. Several criminological theories – inspired by situational criminology theories related to lifestyles, rational choice, psycho-biology, etc. – try to explain how prople act to increase or decrease repeat victimization. The direction of this relationship, however, is often ambiguous, according to the theoretical perspective one adopts. Moreover, the meaning may be different, depending on whether we are talking about property or personal victimization. Thus, we cannot determine unequivocally which theories are correct or incorrect and we must rely only on observed empirical evidence.

In this context, before discussing our results, it is interesting to summarize some postulates of the literature regarding theoretical relationships between victimization and the main variables included in the empirical model.

First, men adopt more risky behaviors (drinking, carring a gun) and have a culture of violent conflict resolution in the case of assault that increase the rate of repeat victimization. On the other hand, adult males are less subject to domestic violence, given the greater physical vulnerability of women. The higher income of white or Asian people in Brazil makes them more attractive targets for criminals. Therefore, the distribution of racial groups in space is not random: white and Asian people usually live in areas that are more exposed to property crimes, but less exposed to crimes against persons (assault). On the other hand, ethnic groups with a higher income and that invest in self-protection are less exposed to the risk of repeat victimization.

Second, young people spend more time on the street and take more risky behaviors in terms of drinking, carrying a gun, etc. Higher testosterone can also increase the risk of victimization by assault, as well as hasty reactions. Young people rely less on police, increasing the chances of impunity for perpetrators of crimes. On the other hand, they tend to travel in groups, reducing their exposure to risk. The physical vulnerability of elderly people can, in turn, expose them to greater risk of victimization.

Third, married people spend more time at home, implying less risky behaviors and more intense surveillance of their homes and other property. Regarding family size, larger families are associated with lower incomes, reducing their attractiveness, but large families increase the surveillance of their property. On the other hand, high density (many residents in the same space) can enhance the risk of domestic conflicts.

Fourth, individuals living in urban areas are more anonymous, have a greater supply of goods and are less exposed to the possibility or victims running into offenders more than once, reducing the risk of detection of crime and punishment.

Fifth, common houses are particularly vulnerable to theft and residential burglary because of their architecture. On the other hand, this perceived vulnerability can encourage their owners to protect or guard them more intensely. Victimized owners increase the level of post-incident protection, reducing their exposure to repeat victimization.

Sixth, with regard to education, students spend more time on the streets (in recreational and social activities) and adopt more risky behaviors. However, travelling in groups and alternative forms of conflict resolution can reduce the risk of victimization by assault. Schooling is associated with a higher income, a fact that makes more schooled individuals more attractive as targets. Also in this case, alternative forms of conflict resolution and improvements in the social environment can reduce the rate of victimization by assault. Individuals who work are more exposed to risks because their homes remain more unprotected when they are out working and their income might be higher, enhancing their attractiveness to criminals. On the other hand, unemployment can increase the risk of assaults if associated with alcohol abuse, spending long hours on the streets, partying, low self-esteem, conflicts in the family. In sum, workers are more exposed to risks as they spend more time on the streets, their homes remain unprotected, their income increases and, as a result, their attractiveness as targets to criminals increases. More working hours also increase the risk of victimization.

Finally, higher-income individuals are more attractive to criminals. The distribution of income groups in space is not random: the affluent live in areas that are more exposed to property crimes but less exposed to crimes against people. On the other hand, higher income people invest more in self-protection. Owning a vehicle is another proxy variable for economic attractiveness, especially for vehicle theft/robbery. Conversely, it reduces the use of public transportation, which can be a factor of protection.

As noted, for almost all variables there are good theoretical reasons to expect both positive and negative effects on victimization. Therefore, the contribution of this paper is that of investigating the net effect of these variables on repeat victimization, which we believe is not different from simple victimization (i.e. a single victimization).

Next, we describe and briefly discuss the main results found. However, before showing the results for repeat victimization models, we analyze how the same variables affect simple victimization. Table A.2 shows the results of the simple victimization model, in which the same regressors were used.

The simple victimizations model suggests that some variables increase the likelihood of victimization: being male, older, living in urban areas, studying, working many hours a week and having higher incomes. In contrast, the sign is negative (i.e. it decreases the likelihood of victimization) for factors such as being white or Asian (for consummated theft only), being married, living in a house, and having a large family.

Regarding assault (crime against a person), it is interesting to note that most variables maintain their significance and sign, but in some cases the sign appears inverted: studying and years of schooling increase the likelihood of property victimization, but for assault the effect is opposite, i.e. it protects potential victims. This demonstrates the need to distinguish the different types of crimes in the analysis, as sugested in Cohen et al. (1981).

At this point, one must remember that the two hypotheses stated in the introduction of this paper were: first, that income has a nonlinear positive effect on repeat victimization; and second, that income has a nonlinear negative effect on repeat victimization. Table Table 2 shows the results of the repeat victimization model.

Our results indicate that, on the one hand, income has a nonlinear positive effect on the amount of times an individual was victimized by property crimes in a one-year period. On the other hand, evidence is provided of a nonlinear negative effect on assault. So the hypotheses tested here are not rejected. This evidence reinforces the economic approach, namely, that income determines victimization based on two factors: economic attractiveness and investment in self-protection. This approach was presented in Section 2. The results also corroborate the sociological thesis according to which wealthier people are less exposed to the risk of suffering aggravated physical aggression such as assault, for instance. One can understand that the results found for the model of victimization by assault can be used as counter-evidence of the robustness of the results found for the three property crimes analyzed in this study. We believe so, particularly because in the case of assault the effect of income was negative, while for other crimes (against property) it was positive.

Regarding other results, the first important point to make is that the variables and sign found for simple victimization remains virtually unchanged when we analyze the repeat victimization model. This piece of evidence reinforces our suspicion that the net effect of the main variables on repeat victimization is the same when compared to simple victimization. Moreover, it corroborates the ideia that it is possible to use the traditional sociological approach proposed by Cohen et al. (1981) to analyze repeat victimization as well. Similarly, the economic approach used in Gaviria and Pagés (2002) is also useful for this purpose. This is an important issue that was clarified in this study.

Regarding the repeat victimization model, it was seen that crimes increase in average if a person is male, older, lives in an urban area, studies or works, is highly educated, works for many hours a week and has a higher income. On the other hand, a person is more protected against repeat victimization if he or she is white or Asian (for consummated theft only), married, lives in a house or has a large family. As in the case of simple victimization, the variables that control for being a student and having more years of schooling increase the chances of repeat victimization, but the sign is inverted in the case of repeat assault.

Others variables also have a positive effect on repeat victimization, namely, being male, young, living in an urban area, studying, having a high level of education, working for long hours, and having a higher income. On the other hand, protective factors that reduce the chances of victimization include being white/Asian (for theft only), being married, living in a house, owning a vehicle and having a large family. We found that for some variables, such as being a student and years of schooling, the effect shifts direction according to the kind of victimization as risk factors turn into protective factors in the case of assault.

Table 2: Marginal effects calculated from repeat victimization models

Variables	Theft		Robbery		Attempted theft/robb.		Assault	
	NB	ZINB	NB	ZINB	NB	ZINB	NB	ZINB
man	.00999*	.0111*	.00932*	.0101*	.0152*	.0162*	.00400*	.00234***
	(.00116)	(.00131)	(.000912)	(.00117)	(.00132)	(.00153)	(.00104)	(.00122)
white	−.00329**	−.00274***	−.000674	−.000490	.000117	.00122	−.0000392	−.00179
	(.00144)	(.00156)	(.000967)	(.00121)	(.00158)	(.00191)	(.00116)	(.00140)
age	.00191*	.00187*	.00113*	.00133*	.00282*	.00291*	.000621*	.000387*
	(.000203)	(.000168)	(.000159)	(.000175)	(.000250)	(.000222)	(.000178)	(.000147)
agesq	−.0000161*	−.0000161*	−.0000132*	−.0000150*	−.0000269*	−.0000274*	−.0000128*	−.0000100*
	(.00000222)	(.00000180)	(.00000187)	(.00000193)	(.00000277)	(.00000236)	(.00000209)	(.00000169)
married	−.00767*	−.00842*	−.0116*	−.0148*	−.0105*	−.0129*	−.0128*	−.0123*
	(.00134)	(.00143)	(.00105)	(.00142)	(.00152)	(.00177)	(.00124)	(.00101)
urban	.0250*	.0271*	.0315*	.0324*	.0417*	.0411*	.0119*	.0143*
	(.00160)	(.00158)	(.00139)	(.00175)	(.00204)	(.00245)	(.00129)	(.00124)
house	−.00666**	−.0133*	−.00903*	−.0202*	−.0202*	−.0345*		
	(.00264)	(.00320)	(.00204)	(.00289)	(.00348)	(.00411)		
student	.00283	.00452*	.00313**	.00274***	.00575*	.00817*	−.00648*	−.00688*
	(.00199)	(.00174)	(.00136)	(.00161)	(.00214)	(.00226)	(.00130)	(.00131)
school	.00129*	.00182*	.00230*	.00343*	.00312*	.00435*	−.000660*	−.000504*
	(.000172)	(.000169)	(.000141)	(.000191)	(.000199)	(.000218)	(.000153)	(.000135)
hwork	.000348*	.000440*	.000336*	.000432*	.000424*	.000545*		
	(.0000315)	(.0000323)	(.0000270)	(.0000282)	(.0000388)	(.0000368)		
works							.00249**	.00380*
							(.00117)	(.00101)
ownveh	.00100	.00209	−.00509*	−.00785*	−.00218	−.00301***	−.00991*	−.00886*
	(.00148)	(.00139)	(.00109)	(.00124)	(.00175)	(.00176)	(.00130)	(.00107)
ln(inc)	.00340*	.00263*	.00438*	.00611*	.00569*	.00744*		
	(.000990)	(.000764)	(.000707)	(.000770)	(.00110)	(.00103)		
famsize	−.00475*	−.00422*	−.00226*	−.00264*	−.00578*	−.00597*		
	(.000509)	(.000419)	(.000390)	(.000398)	(.000646)	(.000604)		
ln(incp)							−.00216*	−.00187*
							(.000722)	(.000646)

Notes: $n = 318,774$; Table A.1 shows the precise description of the variables; dummy variables for states were used; Table A.3 contains the estimated coefficients.

Finally, it must be said that according to the 2009 PNAD survey, the feeling of insecurity in Brazil increases significantly as the per capita household income rises. For example, in the metropolitan area of São Paulo, 42.9% of all individuals with a per capita household income below one-quarter of the minimum wage felt safe. This percentage decreases as the income level increases, to the point of being up to 12.7 percentage points lower for two or more people earning minimum wages. On the other hand, people tend to invest more in self-protection equipment to reduce the risk of victimization when they are victimized more than once, i.e. when they suffer repeat victimization. The questionnaire of the 2009 PNAD survey includes several questions about the use of safety devices for self-protection in one's home. But the response is binary, yes or no, so it is not possible to know how much money the individuals are spending on self-protection. Moreover, this variable is clearly endogenous. We cannot know, for example, if their investiments in self-protection were made before or after their first victimization. For this reason, we chose not to use this variable as a regressor. But, the role of self-protection in repeat victimization was at least partially controlled for when household income was taken into account. However, an advance would consist in finding a set of valid instruments for the use of self-protection means.

5. Concluding remarks

The main objective of this paper was to estimate the effect of income on theft, robbery, attempted theft/robbery, and assault. We found evidence supporting the hypotheses tested that income has a nonlinear positive effect on the number of times an individual was victimized by property crimes, but a nonlinear negative effect on assault – a crime against a person.

When we compared the simple victimization model to the repeat victimization model, no major differences in terms of signs of coefficients were found. In other words, the difference between the two phenomena is not in nature, but in scale. Moreover, we found that it is very important to distinguish the different types of crimes in the modeling.

The results of this paper corroborate the main evidence found by Carvalho and Lavor (2008), although the marginal effects were different because three crimes against property were analyzed separately here.

It must be said that some risk factors can be modified through public policy or individual actions, but others simply cannot. Nobody can change one's color, age or gender to prevent crime. No one will leave school or a job, marry or purposely earn less just to reduce his or her exposure to risk. It is also possible to move or change one's type of residence, but individuals rarely make such decisions for safety reasons alone.

Finally, although we have a better understanding of risk factors associated with victimization and repeat victimization, they will continue to manifest themselves. Those who accumulate "risk factors" will continue to have a higher probability of being victimized than others. There are additional reasons for repeat victimization: a first successful action on a target – mainly on fixed targets such as homes or businesses – encourages criminals to repeat it. Also, if protection mechanisms are damaged and not repaired, targets are exposed to the risk of new attacks. In any case, finding out variables that increase or decrease the risk of exposure is not just an academic curiosity. When a group finds out that it belong to a risk group it becomes more alert. Such target group is thus led to review risk behaviors and take more precautions than others.

References

Becker, G. S., 1968. Crime and punishment: an economic approach. The Journal of Political 76 (2), 169–217.

Cameron, A. C., Trivedi, P. K., march 2009. Microeconometrics using stata, 2nd Edition. Stata Press, College station.

Carvalho, J. R., Lavor, S. C., 2008. Repeat criminal victimization and income inequality in brazil. Economia (Selecta) 9 (4), 87–110.

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 (5), 505–524.

Ehrlich, I., 1973. Participation in illegitimate activities: A theoretical and empirical investigation. Journal of Political Economy 81 (3), 521 – 565.
Gaviria, A., Pagés, C., 2002. Patterns of crime victimization in latin american cities. Journal of Development Economics 67 (1), 181–203.
Justus, M., Kassouf, A. L., 2013. Evidence of the effect of wealth observed by criminals on the risk of becoming a victim of property crimes. EconomiA 14 (2), 88 – 101.
Kleiber, C., Zeileis, A., 2008. Applied Econometrics with R. Springer, New York.
Skinner, C., Holt, D., Smith, T., 1998. Analysis of Complex Surveys. John Wiley & Sons, Chichester.

AppendixA.

Table A.1: Definition and summary statistics of the variables

Variable	Definition	Mean	Linearized Std. Error
Responses			
`theft`	Counts of victimizations by theft.	0.0548	0.0011
`robbery`	Counts of victimizations by robbery.	0.0483	0.0008
`attempted theft/robb.`	Counts of victimizations by attempted theft/robbery.	0.0761	0.0013
`assault`	Counts of victimizations by assault.	0.0027	0.0009
Explanatory			
`man`	1 if man and 0 if women.	0.4837	0.0008
`white`	1 if white or yellow (Asian) and 0 if black, mulatto or indigenous people.	0.4893	0.0027
`age`	Age in years.	360.86	0.0578
`married`	1 if married and 0 otherwise.	0.4127	0.0019
`urban`	1 if one lives in an urban area and 0 if one lives in a rural area.	0.8425	0.0047
`house`	1 if one lives in a house and 0 if one lives in an apartment.	0.9141	0.0021
`student`	1 if one studies and 0 otherwise.	0.2357	0.0010
`school`	Years of schooling.	70.14	0.0261
`hwork`	Hours of work in a week.	210.59	0.0634
`works`	1 if employed and 0 if unemployed or out of the labor force.	0.5736	0.0015
`ownveh`	1 if one owns a car or a motorcycle and 0 otherwise.	0.4968	0.0028
`ln(inc)`	Logarithm for monthly household income.	70.22	0.0062
`famsize`	Number of individuals in the family.	30.63	0.0081
`ln(incp)`	Logarithm for monthly per capita household income.	60.02	0.0070

$n = 318,774$.

Table A.2: Results of victimization models

Variables	Theft		Robbery		Theft/robbery attempt		Assault	
	Coef.	Marg. Eff.	Coef.	Marg. Eff.	Coef.	Marg. Eff.	Coef.	Marg. Eff.
`man`	.105*	.00804*	.102*	.00668*	.0927*	.00895*	.148*	.00493*
	(.00977)	(.000745)	(.00972)	(.000645)	(.00860)	(.000839)	(.0129)	(.000439)
`white`	−.0338*	−.00259*	−.0136	−.000890	.00529	.000509	−.0247	−.000812
	(.0113)	(.000860)	(.0107)	(.000699)	(.0103)	(.000996)	(.0154)	(.000505)
`age`	.0163*	.00125*	.0112*	.000733*	.0182*	.00175*	.0111*	.000365*
	(.00159)	(.000121)	(.00166)	(.000108)	(.00143)	(.000140)	(.00243)	(.0000795)
`agesq`	−.000146*	−.0000112*	−.000146*	−.00000956*	−.000183*	−.0000176*	−.000224*	−.00000736*
	(.0000173)	(.00000133)	(.0000192)	(.00000126)	(.0000159)	(.00000155)	(.0000289)	(.000000942)
`married`	−.0773*	−.00584*	−.145*	−.00924*	−.0831*	−.00791*	−.214*	−.00680*
	(.0111)	(.000819)	(.0123)	(.000771)	(.0105)	(.000989)	(.0165)	(.000514)
`urban`	.295*	.0188*	.506*	.0242*	.375*	.0292*	.289*	.00770*
	(.0193)	(.00103)	(.0344)	(.00105)	(.0239)	(.00146)	(.0294)	(.000614)
`house`	−.0655*	−.00526*	−.114*	−.00811*	−.134*	−.0142*		
	(.0195)	(.00163)	(.0190)	(.00147)	(.0184)	(.00211)		
`student`	.0290***	.00225***	.0401**	.00267**	.0476*	.00468*	−.116*	−.00356*
	(.0159)	(.00124)	(.0156)	(.00106)	(.0141)	(.00142)	(.0228)	(.000657)
`school`	.0133*	.00102*	.0293*	.00191*	.0237*	.00228*	−.00792*	−.000261*
	(.00142)	(.000107)	(.00151)	(.000101)	(.00132)	(.000126)	(.00203)	(.0000667)
`hwork`	.00348*	.000266*	.00395*	.000258*	.00321*	.000310*		
	(.000275)	(.0000210)	(.000268)	(.0000179)	(.000232)	(.0000225)		
`works`							.0238	.000781
							(.0163)	(.000532)
`ownveh`	.0224***	.00172***	−.0783*	−.00511*	−.0122	−.00117	−.140*	−.00463*
	(.0119)	(.000908)	(.0120)	(.000784)	(.0114)	(.00109)	(.0162)	(.000538)
`ln(inc)`	.0324*	.00248*	.0628*	.00410*	.0496*	.00478*		
	(.00721)	(.000553)	(.00785)	(.000514)	(.00695)	(.000668)		
`famsize`	−.0458*	−.00350*	−.0280*	−.00183*	−.0461*	−.00444*		
	(.00406)	(.000311)	(.00437)	(.000286)	(.00421)	(.000402)		
`ln(incp)`							−.0379*	−.00125*
							(.0106)	(.000349)

Notes: $n = 318,774$; Table A.1 shows the precise description of the variables; dummy variables for states were used.

Table A.3: Coefficients estimated from repeat victimization models

Variables	Theft		Robbery		Theft/robbery attempt		Assault	
	Nb	Zinb	Nb	Zinb	Nb	Zinb	Nb	Zinb
man	.214*	.209*	.259*	.231*	.244*	.218*	.193*	.0989***
	(.0253)	(.0249)	(.0255)	(.0263)	(.0210)	(.0202)	(.0491)	(.0517)
white	−.0711**	−.0519***	−.0189	−.0112	.00189	.0165	−.00189	−.0757
	(.0311)	(.0296)	(.0271)	(.0277)	(.0255)	(.0258)	(.0563)	(.0596)
age	.0411*	−.0483*	.0315*	−.0332*	.0454*	−.0495*	.0300*	−.0127*
	(.00438)	(.00501)	(.00445)	(.00463)	(.00386)	(.00407)	(.00866)	(.00482)
agesq	−.000348*	.000417*	−.000371*	.000376*	−.000433*	.000466*	−.000619*	.000330*
	(.0000477)	(.0000512)	(.0000524)	(.0000498)	(.0000433)	(.0000420)	(.000103)	(.0000547)
married	−.168*	.218*	−.332*	.375*	−.172*	.221*	−.641*	.415*
	(.0302)	(.0377)	(.0306)	(.0371)	(.0252)	(.0306)	(.0626)	(.0328)
urban	.665*	−.718*	1.279*	−.925*	.877*	−.737*	.719*	−.548*
	(.0521)	(.0506)	(.0933)	(.0646)	(.0578)	(.0491)	(.0975)	(.0545)
house	−.136*	.355*	−.230*	.501*	−.287*	.615*		
	(.0510)	(.0940)	(.0476)	(.0774)	(.0441)	(.0765)		
Student	.0600	−.117*	.0857**	−.0683***	.0904*	−.139*	−.341*	.237*
	(.0417)	(.0454)	(.0363)	(.0399)	(.0327)	(.0387)	(.0734)	(.0460)
school	.0278*	−.0471*	.0644*	−.0858*	.0503*	−.0741*	−.0319*	.0166*
	(.00378)	(.00528)	(.00381)	(.00656)	(.00324)	(.00474)	(.00745)	(.00429)
hwork	.00750*	−.0114*	.00941*	−.0108*	.00681*	−.00927*		
	(.000685)	(.000851)	(.000743)	(.000759)	(.000620)	(.000683)		
works							.121**	−.126*
							(.0572)	(.0334)
ownveh	.0216	−.0541	−.143*	.197*	−.0351	.0512***	−.475*	.292*
	(.0319)	(.0357)	(.0307)	(.0305)	(.0281)	(.0301)	(.0616)	(.0365)
ln(inc)	.0733*	−.0678*	.123*	−.153*	.0915*	−.127*		
	(.0214)	(.0193)	(.0195)	(.0197)	(.0180)	(.0177)		
famsize	−.102*	.109*	−.0633*	.0660*	−.0929*	.102*		
	(.0111)	(.0104)	(.0108)	(.0101)	(.0105)	(.0101)		
ln(incp)							−.104*	.0617*
							(.0346)	(.0204)
Constant	−5.135*	−2.533*	−5.696*	−2.180*	−4.814*	2.275*	−3.579*	.272***
	(.204)	(.0878)	(.201)	(.0874)	(.174)	(.168)	(.351)	(.139)
Const. inflate		1.884*		2.556*		−1.817*		−2.731*
$\ln \alpha$	2.508*	1.956*	2.249*	1.573*	2.186*	1.574*	3.933*	2.786*
	(.0311)	(.0584)	(.0362)	(.0694)	(.0291)	(.0540)	(.0455)	(.0740)
Voung test (ZINB vs. NB)	$z = 28.18$		$z = 38.59$		$z = 38.60$		$z = 17.92$	
	p-value $= 0.000$		p-value $= 0.000$		p-value $= 0.000$		p-value $= 0.000$	

Notes: $n = 318,774$; Table A.1 shows the precise description of the variables; dummy variables for states were used.