Repeat Criminal Victimization and Income Inequality In Brazil

José R. Carvalho* Sylvia C. Lavor†

Abstract

This paper is concerned with the analysis of the impact that socioeconomic characteristics of the individuals and regions, with special emphasis on income inequality, has in the process of repeat victimization. To accomplish this task, a series of count data models are estimated. The zero inflated negative binomial model performs best, in a series of nested and non-nested inferencial procedures. These econometrics models are applied to a unique data set: the 1988 PNAD. For this specific year only, an additional set of questions provide us with a national sample of victimization, something yet not available again until now. Our results are in close accordance to the international literature, and the estimative concerning the impact of income inequality on repeat victimization is quite strong and robust to the use of different measures. The elasticity of the GINI coefficient is positive and higher than one (1.3012), meaning that increasing inequality leads to more victimizations. The results about the effect of inequality has the potential to offer new insights for security policies at both the federal and state level in Brazil.

Key words: Victimization, Inequality and Count Models

Resumo

Este artigo se propõe a analisar o impacto das características socioeconômicas do indivíduo e da região, com ênfase sobre a desigualdade de renda, no processo da vitimização repetida. Para atingir esse objetivo, modelos de dados de contagem são estimados. O modelo de zeros inflados negativo binomial apresentou melhores resultados, em uma série de procedimentos inferenciais aninhados e não aninhados. Esses modelos econométricos foram aplicados para uma base de dados singular: a PNAD de 1988. Para este ano específico, uma série de questões adicionais forneceu uma amostra nacional sobre a vitimização, algo até então não disponível. Nossos resultados se mostraram de acordo com a literatura internacional e a estimativa do impacto da desigualdade de renda sobre a vitimização foi bastante forte e robusta ao uso de diferentes medidas. A elasticidade do coeficiente da variável Gini foi positivo e maior do que um (1,3012), significando que um aumento na desigualdade leva a mais vitimização. Diante disso, os resultados obtidos acerca do efeito da desigualdade oferecem novas leituras para as políticas de segurança tanto em nível federal quanto estadual no Brasil.

Palavras-chave: Vitimização, Desigualdade e Modelos de Contagem
JEL: K42, A13 e C25
ANPEC: Área 11 - Economia Social e Demografia Econômica

*CAEN/UFC and Laboratório de Econometria - CAEN/UFC, Brazil. Author for correspondence (josecarv@ufc.br).
†Laboratório de Econometria - CAEN/UFC, Brazil.

1 Introduction

During the last years violence and crime has become one of the most prominent issue in the media and all over the population in general. More specifically, crime is seen as a restriction to the economic and social development of a region and, as a consequence, formulating effective policies to fight against crime appears to be a question with high social and economic returns (see, Barslund, Rand, Tarp, and Chiconela (2005)).

Araújo and Fajnzylber (2001) consider that the phenomenon of criminality consists of a problem that presents diverse facets: it is social, because it affects directly the quality and the life expectancy of populations; it is economic, because it limits the potential development of economies; and it is political, since the necessary actions for fighting against crime involve the active participation of governments and the respective allocation of public resources. With significant increases in the rates of criminality and the growing importance given to the subject, governments started to face a challenge to formulate and to implement policies that can prevent and, consequently, reduce crime. To help the elaboration of these policies, it is of utmost importance the development of studies that allows the understanding of the determinants of crime activity.

Victimization, a closely related concept to crime, has not yet received the necessary attention from schoolars, however. Victimization has long been recognized by criminologists as an important perspective to understand crime. The appearance of studies on the subject of victimization made possible to analyze the socioeconomic characteristics of the individual and/or the family who suffered any type of delict and, also, to verify the determinants that influence in the repeated victimization process. In other words, these studies had favored the identification of characteristics that contribute to most individual victimization. Some empirical studies had shown that is important to make research on victimization, mainly focusing on the reasons that conduce to it repeated incidence. According to Laycock (2001), it is possible to create and to implement programs to help fighting against recurrent victimization, that can be able to build a strategy for crime reduction. A key issue that received a great deal of attention is the impact of income inequality on victimization.

Some studies, such as Fajnzylber and et al. (1998), make use of inequality income indicators to show that the greater the concentration of income in a determined place, the greater will be the incidence of delicts. Also, papers as Levitt (1999), using data on the average income of the family, indicate that, in contrast of what people might imagine, the poorer people have greater chances of suffering with victimization than the richer people do. Although there exists a great deal of discussion about this subject, there are few papers that study victimization at the individual level, in other words, that take in consideration the role played by the victim at the moment of the delict. The majority of studies analyzes crime under the perspective of the delinquent. However, it is necessary to make clear that there exist two ways to study crime: focusing on the side of the criminal, studying the reasons why he/she has committed the crime; or analyzing the side of the victim, verifying which characteristics he/she possess that can influence his/her own victimization.

In face of this fact, this paper intends to analyze, in a more inclusive form, the victimization process in Brazil, giving emphasis to the characteristics of repeat victimization. In other words, a brief analysis of the characteristics of the individual that became a victim is made, independent of the amount of times that the event manifested, and which factors are determinant so that the same one comes to suffer from repeat victimization. At the same time, an analysis on the relation between inequality of income and victimization in the country is performed, addressing the question whether income concentration affects the incidence of crimes and the behavior of the victims. For the accomplishment of this task, the data of the national Research for Sample of Domiciles (PNAD) of 1988, a unique occasion that a data collection with a national character approached questions related to the subject of the victimization, is used. Although it is not officially a survey on victimization, the sampled data made possible an analysis regarding victimization in Brazil. To complement this database, information on economic variables through the Institute of Applied Economic Research (IPEADATA) for the same year, had been used.

By means of a sequence of count data econometric models, we are able to replicate international results, as well as get new ones, concerning the positive effects[1] of variables such as urban local of residence, being the head of the household and status of employed. Race and sex, surprisingly, have no statistically significant effects. The guardianship effect, where households with a larger number of components have lower mean number of victimizations, is found in our data set. The effect on income is positive, prompting us to conclude that the effect of attractiveness of the target dominates the effect of private protection.

The key issue under analysis, i.e., the effect of income inequality, proxied by the Gini index, on repeat victimization bring us surprising new evidence. The effect is positive and turns to be the largest among all independent variables (1.3012)! Actually, this result is robust to the use of different measures such as the Theil index. As far as we know, our paper is the first to obtain this result using a methodology based on count data models. Even though our results must be taken as not definitive, they point out to some important issues that must be considered in the formulation of policies of development and security so debated among political and academic spheres in Brazil.

The structure of the paper is comprised, besides this introductory section, of more 5 sections. Section 2 makes a literature review about repeat victimization and income inequality. Section 3 performs a detailed analysis of the available empirical evidence on repeat victimization, using the 1988 PNAD as the main source. Section 4, in a rather detailed way, develops the main count data models that are used for estimation. Section 5 presents the estimated parameters as well as perform the necessary inferential procedures. Section 6 concludes the paper, offering some avenues of future research.

2 Literature Review

The majority of studies that currently exists in the literature on crime had its initial motivation in Becker (1968), who shown that the relation between economic incentives and criminality is a consequence of rational behavior. In other words, Becker suggested that the criminal actions can directly be seen as connected to the rational decisions of the criminal, where he/she observes the costs/benefits of the crime commitment.

Although Becker's original work prompted a huge amount of research, with many other empirical studies connected to the assumption of rational choice, the criminological literature has pointed out the importance of paying attention to the victimization process. The main point about the victimization process is that a crime does not necessarily turns out to be recorded as a crime. In fact, as it is well known to criminologist and sociologist but unfortunately less well known among economists, crime under-reporting is far from being just a curiosity with no systematic explanation. The "dark figure", i.e., victimizations that are not officially reported, is a well established regularity. Among the very few economic papers that recognized this important fact are MacDonald (2001) and MacDonald (2002).

Before we go into details on some papers, it is important to digress a little bit about the concept of repeat victimization. The concept is very easy to understand. Given a specific time range, the larger this period the higher the probability of one experiencing one or more episodes of victimization. This

[1]Positive effect means that an increase in the independent variable leads to an increase in the mean number of repeat victimizations

issue was incorporated, with great success into the criminological main stream by Pease (1998). By noting that the best predictor of victimization is past victimization, the author developed an approach of criminal policy based on the notion of repeat victimization. In fact, now is well known that the amount of victimization is heavily concentrated among few people (see, Pease (1998), Farrell and Pease (2001) and Tseloni and Pease (2003), among others).

The literature on repeat victimization is now huge[2], however is almost entirely located among criminologists and few sociologists. Repeat victimization measures are entirely neglected by econometricians and economists. Even though discussing the perils of time aggregation of the number of victimizations under a single indicator of victimization occurrence is beyond the scope of our paper, it is important to stress some possible bias of this approach. First, there is the problem of completely missing the fact that victimization is concentrated among certain types of individuals. Second, the number of victimizations in a time interval carries along many behavioral insights. Finally, count models can be rationalized by assuming the existence of heterogeneous groups of people: one that never experience a victimization and the other that are comprised of people who can experience any amount of victimization, including zero. By aggregating repeat victimization we loose the appeal of applying such count data models. So, we are going to treat the victimization process as one of repeat victimization.

As mentioned before, our approach take a different view by using victimization as the key dependent variable in a model of crime[3]. This is a growing tendency among economists and econometricians. So in order to study the impact of income inequality on crime, we make a point in favor of using (repeat) victimization as the best choice. We are not alone in this regard. The paper of Gaviria and Pagés (1999) analyzed the relation between victimization and income classes. They employed a data set from 17 countries from Latin America for the years 1996 and 1998. They find that the majority of property theft come from the high and middle classes who live in larger cities. Besides that, city growth is correlated with victimization rate. However, they use the Latinobarometer data set (Latinobarometer (1999)) which falls short as being a good victimization survey.

For a more geographically focused study, it is worth mentioning Tella, Galiani, and Schargrodsky (2002). These authors try to asses the correlation between income distribution and victimization. The key research question is to figure out which social group suffered the most (in terms of higher victimization rates) with the growing rate of criminality in Argentina. They used a survey questionnaire to collect information on victimization at metropolitan Buenos Aires. For household victimization, poor people suffered more with increasing criminality. They try to explain this fact by arguing that richer people are able to invest in private security, something less likely for poor people. Also, the very private spending on private security from the part of rich people creates negative externalities[4] for poor people. Although their results are robust, they do not consider repeat victimization.

Another important paper is Bourguignon, Sanchez, and Nunez (2003). Even though they do not use victimization, their careful analysis of the impact of income inequality and crime is illuminating. The authors develops a "structural" model of crime. Their main point is to rationalize some negative results regarding the effect of inequality on crime. They argue that usual measures of inequality, such as Gini and Theil indexes, fail to represent the inequality change that really matters for criminal activity: those changes located in a specific income range, close to the bottom of the distribution. After developing a clever model, they estimate it with Colombian data and find the "right" effect, i.e., the higher the inequality the greater the crime rate. Note that, as we will argue later on, it might be possible to rationalize their result by saying that crime should be measured by victimization. It means

[2]For a recent paper and account of the issue, see Tseloni and Pease (2004)

[3]Actually, as we will make clear, we use repeat victimization measures.

[4]By displacing criminals to poor areas.

that it is likely that some existing "puzzles" in the crime literature might be due to the economists lack of knowledge about the difference between crime and victimization.

Finally, Barslund, Rand, Tarp, and Chiconela (2005) employ a micro data set from Mozambique to study the determinants of victimization. Among other correlates, they include income. Their main results point out to a robust positive effect of income on victimization, although at decreasing rates. Inequality, as well as unemployment have the effect of increasing victimization rates. Interesting, their results show that even though the poor is less likely to suffer thefts[5] they suffer more, relatively to rich people, as they have a lower chance of recomposing the lost what has been taken. Despite the revealing results, they restrict themselves to victimization, disregarding the more important issue of repeat victimization.

Another stream of research related to ours started with Levitt (1999). This seminal paper initiated an approach to studing victimization that has as the main focus a preoccupation with questions regarding crime distribution and inequality in income distribution. Researchers who study this issue have found constantly that crimes are distributed differently among social groups[6]. Therefore, the central question to be examined consists of whether an increase in inequality of a region' income impacts in the victimization and which social groups suffer more with this.

Using average income as inequality indicator, Levitt (1999) asserts that due to the income inequality increase in the United States, the less-favored people have started to suffer more with victimization, because crimes started to be more concentrated in the poor areas. A possible explanation for the reduction of the victimization risk among rich people, according to the author, could be the increase of private security expenses. In other words, with greater purchasing power, one can invest more in this type of security. Beyond investigating victimization among social groups, Levitt (1999) also analyzes this phenomenon along the race dimension. According to the author, not whites who has a determined income are victimized more frequently than whites. Moreover, the former are twice more vulnerable to be robbery victims than the later; and those who are not white or rich, get twice as more chances to have their vehicle stolen than the rich whites.

An important question about the difference in the victimization distribution among social groups is that income level influences the way that a richer person protect herself from some delict. As discussed previously, well-taken care with private security are recognized by empirical studies as capable to reduce the level of crime in some areas, but can increase in others, because this type of protection can generate negative externalities in poor areas.

Another paper that approaches the relation between income inequality and victimization is Thacher (2004). The author shows that despite the decline in criminality during the period from 1974 to 2000, in the United States, this benefit was not shared equally. Said in other words, poor people have started to suffer more with victimization than people with higher income in all the types of delict, confirming the results gotten in other studies, in which increasing the inequality of income distribution, impacts more individuals with lower purchase power in terms of victimization.

In Brazil, most of the debates on this issue refer to the influence of socioeconomic factors on criminality, beyond the measurement of the cost of violence for society. However, the available studies that investigate the problematic around criminality are limited. Among these few, none verifies the relation between income inequality and criminality. Regarding studies about the link between victimization and inequality we found only one paper in Brazil. Gomes and Paz (2007) builds a "structural"

[5]Of course, this is a reasonable fact.

[6]Note that such approach can be found in the initials writings of Becker. In fact, in his work it has been recognized that a rich person becomes a more attractive target when her income increases. However, if the rich behaves in such a way that it reduces his risk of victimization, by investing in private security, then his risk of victimization can actually diminishes. Consequently, the poor' victimization can increase in relation to the rich one.

model of victimization that serves as a basis for the estimation of a binomial model of victimization for the population of São Paulo, the largest brazilian state. Even though their main focus is on the city size effect on victimization, they are able to find an increasing non-monotonic income effect on victimization. However, they do not address inequality in income issues, nor they include an analysis of multiple victimization.

In face of this literature review, we believe we identified some important gaps, both nationally and internationally, in the literature of victimization and income inequality that are worth pursuing, such as:

1. Very few international studies employ repeat victimization measures and no national study does it;

2. The data set used are generally comprised of small samples[7].

3. None paper in Brazil include inequality measures as a possible predictor of victimization.

Hence, in face of these gaps, this work will focus, from now, on the repeat victimization process in Brazil, verifying which individual characteristics can impact on victimization occurrence and whether income inequality influences this phenomenon. The data set used, as mentioned, is the 1988 PNAD[8]. The econometric model is based on the estimation of count data regressions that will try to control for both unobserved heterogeneity and excess of zeros present in the sample. Next section will address the empirical evidence.

3 Empirical Evidence on Victimization

Recently, some empirical studies have shown great interest in repeat victimization issues. In the last two decades, it was recognized that repeat victimization of people and/or places represents the majority of all occurred victimization in some regions, e.g., Kleemans (2001). Moreover, Sparks (1981) showed that this issue occurs more frequently than it could foresee.

In accordance with Ybarra and Lohr (2002), there are many evidences that victimized people have a higher probability of suffering future victimizations than those ones that had never been victims. Such evidence can be observed by following individual behavior after the occurrence of a delict. Moreover, the study of Kleemans (2001) reveals that repeated victimization is more likely to occur in areas with high criminality rates, suggesting that crime concentration is linked with the number of victimizations.

In face of this evidence, it can be said that the odds of a person becoming a victim, considering that he/she has already suffered some type of delict, are increasing. Said this, this section analyzes the available empirical evidence using as the main source of information on victimization: the 1988 PNAD. This research is the most important household survey employed in Brazil. Besides the data contained in PNAD, we will use other empirical evidence about the economic characteristics of the region, for instance, Gini index of each state, from the IPEADATA, for the same year and other measures of inequality and poverty. Even though that data set is not a victimization oriented survey, in 1988 it

[7]With the exception of american and britsh studies that use the Criminal Victimization Survey and the British Victimization Survey, respectively. These data sets are large and well designed surveys specifically developed to track the pattern of (repeat) victimization

[8]Admittedly this is a rather old data set! However, if one considers both the richness of the sample, its national coverage as well as the fact that there is no national victimization survey in Brazil as to July, 2008, the 1988 PNAD has a lot of appeal. As a matter of fact, the brazilian government is planning its first national victimization survey for the second semester of 2008.

Table 1: Frequency of Attempts and Actual Victimizations

Count	**Attempt (%)**	**Victim (%)**
$P(Y=0)$	98.00	93.76
$P(Y=1)$	1.57	5.04
$P(Y=2)$	0.22	0.73
$P(Y=3)$	0.12	0.27
$P(Y=4)$	0.02	0.08
$P(Y=5)$	0.01	0.04
$P(Y=6)$	0.006	0.01
$P(Y=7)$	0.00009	0.006
$P(Y=8)$	0.002	0.01
$P(Y=9)$	0.00009	0.00
$P(Y \geq 10)$	0.008	0.01
Total	100.00	100.00

Table 2: Frequency of Victimizations

Count	**Victim (%)**
$P(Y=0)$	92.62
$P(Y=1)$	5.32
$P(Y=2)$	1.31
$P(Y=3)$	0.37
$P(Y=4)$	0.17
$P(Y=5)$	0.07
$P(Y=6)$	0.04
$P(Y=7)$	0.01
$P(Y=8)$	0.01
$P(Y=9)$	0.002
$P(Y \geq 10)$	0.02
Total	100.00

brought as a special feature a small set of questions amenable to be used in a victimization study. The exact wording of the questions were:

1. ''During the period of october/1987 - september/1988 have you ever been a victim of attempted burglar or attempted theft?"

2. ''During the period of october/1987 - september/1988 have you ever been a victim of burglar or theft?"

Table 1 shows the unconditional distributions for attempt and victimization for theft and burglar. However, for the sake of simplicity and because we lack a theory for attempted victimization, we aggregate attempts and actual victimizations under the common denomination of victimization and show its distribution on Table 2.

A feature of victimization data sets is the fact that the conditional distribution of having two or more episodes of victimization is increasing. This implies that the greater past victimization experiences the person has suffered, the higher the likelihood of becoming a victim again. This result coincides with the same ones gotten from Ybarra and Lohr (2002) and Kleemans (2001). Table 3 clearly shows this pattern.

Table 3: Conditional Victimization

Probability	**Result (%)**
$P(Y=1)$	5.75
$P(Y=2\|Y=1)$	19.74
$P(Y=3\|Y=2)$	22.21
$P(Y=4\|Y=3)$	32.36
$P(Y=5\|Y=4)$	29.61
$P(Y=6\|Y=5)$	39.13
$P(Y=7\|Y=6)$	20.16
$P(Y=8\|Y=7)$	56.89
$P(Y=9\|Y=8)$	15.38
$P(Y\geq 10\|Y=9)$	90.62

For our purposes, questions related to personal victimization will be analyzed, this means that will be given an approach to the question of whether the individual was a victim of robbery or theft during the period from October of 1987 to September of 1988, and how many times this occurred. The following table shows the observed distribution of repeat victimization for that period. In accordance with table 2, the distribution is overdispersed, with variance higher than the mean (Mean=0.11 and Var=0.28). Moreover, this sample is composed of a high proportion of zeros. So, negative binomial and zero inflated models seems to be the most appropriate to model repeated victimization. The explanation for this choice will be better displayed in the section of Methodology.

The model to be estimated possesses as dependent variable the amount of times that an individual was victim of robbery or theft, assuming discrete values bigger or equal to zero. Moreover, the model has as independent variables the socioeconomic characteristics of the individual, such as where he lives, sex, color, if he is or not the family head, if he worked in the same week that he was interviewed, age, scholarly, number of components of his family, per capita familiar monthly income and the percentage of poor households, beyond information on income inequality, proxied by the index of Gini and the ratio between the 10% richest income and the 40% poorest one. In this way, the estimated model has the following form:

$$Y = X'\beta + \epsilon \tag{1}$$

where Y is the dependent variable, X is the vector with the independent variables and β the parameters to be estimated. In the next table, one brief description of the variable used in the model that had to be estimated. Moreover, the related table shows the main descriptive statistics.

By inspection of table 4, it is already possible to make a preliminar analysis on the variables. Accordingly, great part of the sample is found in the urban zone. Moreover, a little more than 50% of the sample is formed by women and people who considered themselves to be of white color. More than half of the sample consists of people who said that they are not heads of the family in the household where they live and that they work in some activity. Another point to be detached is that the average age of the interviewed people is 30 years and that the average time of studies for them is a little more than four years.

Table 4: Description of the dependent and independent variable

Variable	Description	Min	Max	Mean	Var
Y	Counts of victimization	0	40	0.11	0.282
SIT	Dummy for localization (Rural = 0 and Urban = 1)	0	1	0.78	0.170
SEX	Dummy for Sex (Female = 0 and Male = 1)	0	1	0.48	0.250
RACE	Dummy for race (Non-white = 0 and White = 1)	0	1	0.52	0.250
CONDD	Dummy for Condition in the household (Non head = 0 and Head = 1)	0	1	0.33	0.221
AGE	Age in years	12	109	33.72	288.770
ACTIVITY	Dummy for employment status (Unemployed or out of the labor force = 0 and Employed = 1)	0	1	0.58	0.244
EDUCATION	Years of education	0	10	4.51	10.738
NUMCOMP	Family size	1	17	4.81	5.151
INCOME	Per capita familiar income (in 1,000 of Cz$)	0	4500	33.36	4666.251
GINI	Gini index	0.4812	0.6573	0.5947	0.001
POOR	% of Poor households	0.15	0.79	0.41	0.031

Regarding to the inequality characteristics, Gini index is, on average, 0,5949 and, on average, 41% of Brazilian households are considered poor. Next section develops and estimate two models: the Negative Binomial Model and the Zero Inflated Negative Binomial.

4 Econometric Methodology

Models of counting data are appropriate to estimate the number of occurrences of events. In this case, the dependent variable Y assumes non-negative integers $(0, 1, 2, \cdots)$, that correspond to the number of events realized in a determined period of time. For instance, events should be: number of visits to a doctor that a person makes during the month; number of times that someone was assaulted in the last year, number of purchased items in a supermarket and so on. Hence, for our analysis, i.e., repeat victimization, count models should be the most appropriate approach. Before developing the econometric model it is worth digressing a little bit about the different types of count models.

A straightforward, but usually incorrect approach, is to apply linear regression directly. Even though there exists situations in which the linear regression model offers reasonable results, is much safer to use models specifically designed to deal with count data. Some models under that label are: Poisson Regression Model (PRM), Negative Binomial Regression (NBR) and variations of these models, as the Zero Inflated Poisson (ZIP) and the Zero Inflated Negative Binomial (ZINB) regression. Generally, the PR model is the starting point for analysis, even though its use could be so frequently inadequate (see, Cameron and Trivedi (2005)).

Let Y be a variable that indicates the number of times that an event occurs along a specified time interval. If Y has a Poisson distribution, then:

$$Pr(Y|\mu) = \frac{e^{-\mu}\mu^{Y}}{Y!} \quad for \quad Y = 0, 1, ... \tag{2}$$

A key fact about this model is the property that $E(Y) = Var(Y)$. Note also that $\mu > 0$ is the only parameter defined in the distribution. The Poisson regression model (PRM) is derived from the distribution of Poisson for the parameterization of the relationship between the mean μ and a vector of regressors X. The standard hypothesis is to use the following structural equation:

$$\mu_i = E(Y_i|X_i) = \exp(X_i'\beta) \tag{3}$$

Again, the mean is equal to the variance, i.e., $\exp(X_i'\beta) = Var(Yi|Xi) = \mu_i$, meaning that the Poisson regression is intrinsically heteroskedastic.

However, our sample data presents different values for mean and variance, which goes against the property of the Poisson distribution. Such kind of pattern is typical of data sets on repeat victimization. So, following the proposal of Tseloni and Pease (2003), we need to find a better specification to deal with overdispersion. The negative binomial regression model (NBR) corrects the imperfection of the PRM by adding a parameter δ_i that reflects the heterogeneity unobserved between the observations. The NBR adds an error ε that is assumed to be uncorrelated with de components of the vector X.

$$\widetilde{\mu_i} = \exp(X_i'\beta)\delta_i \tag{4}$$

Where $\delta_i = \exp(\varepsilon_i)$. To identify the model, it is assumed that $E(\delta_i) = 1$, which corresponds to $E(\varepsilon_i) = 0$ in the linear regression model. It is straightforward to see that now we have $E(\widetilde{\mu_i}) = \mu_i E(\delta_i) = \mu_i$, i.e., the NBR model has the same mean as the Poisson regression model, however its variance allows for over-dispersion. We model overdispersion due to the heterogeneity unobserved between the individuals. The probability distribution for the NBR model must be conditioned on both the values of X such as on δ_i:

$$Pr(Y_i|X_i, \delta_i) = \frac{\exp^{-\widetilde{\mu_i}} \widetilde{\mu_i}^{Y_i}}{Y_i!} \tag{5}$$

Since δ_i is unknown, one cannot calculate $Pr(Y|X)$. However, this could be overcome by assuming that the uncertainty about δ_i can be represented by a gamma distribution. Then, $Pr(Y|x)$ can be calculated as a weighed combination of $Pr(Y|X, \delta_i)$ for all the values of δ_i, where the weights are determined by $Pr(\delta_i)$. So, the binomial negative distribution is given by:

$$Pr(Y|x) = \frac{\Gamma(Y - \alpha^{-1})}{Y!\Gamma(\alpha^{-1})}(\frac{\alpha^{-1}}{\alpha^{-1} + \mu})^{\alpha^{-1}}(\frac{\mu}{\alpha^{-1} + \mu})^{Y} \tag{6}$$

where Γ is a function gamma. Despite of the frequent use of these models, it is common to find counting data sets presenting excess of zeros[9]. In this situation, the application of these models can generate inconsistent estimates. In order to deal with this issue, we decided to use the zero inflated model.

For zero inflated models, the dependent variable Y_i assumes again non-negative integer values. However, there is a crucial difference in those models. There are two unobserved sub-populations:

[9]In the repeat victimization literature, the observed proportion of zeros is frequently beyond 85 % .

those who will never experience any count and those who could experience any number of counts, including zero. Let define the first sub-population by $A = 0$ and the other sub-population by $A = 1$. Since by assumption group membership is not observed, one can model group membership by, say, a logit equation:

$$Pr(A_i = 0|X_i) = F(X_i'\theta) \tag{7}$$

Where F is the logit or probit function. For those belonging to the sub-population where $A = 1$ it is assumed that conditional on the event that $A = 1$, the probability distribution of the number of counts follows a specific count model. For instance, if we assume a Poisson model, the outcome will be the Zero Inflated Poisson Model (ZIP), otherwise if we assume a negative binomial model, it follows the Zero Inflated Negative Binomial Model (ZINB). Note that technically the zero inflated models need to solve a mixing problem: there are two sub-groups, one with a specific marginal probability, i.e., $Pr(A_i = 0|X_i)$, and the conditional distribution of counts $Pr(Y_i = j|X_i, A_i)$, for $j = 0, 1, 2, \ldots$. The observed part of the model comes from the following mixing operation:

$$\begin{aligned} Pr(Y_i = j|X_i) =& Pr(Y_i = j|X_i, A_i = 0)Pr(A_i = 0|X_i) \\ &+ Pr(Y_i = j|X_i, A_i = 1)Pr(A_i = 1|X_i) \end{aligned} \tag{8}$$

The probability of the excess of zeros is denoted by $\pi_i = Pr(A_i = 0|X_i)$, where $0 \leq \pi_i \leq 1$. The mean and the variance of the ZIP random variable is given by:

$$E(Y_i|X_i) = (1 - \pi_i)\lambda_i \tag{9}$$

$$Var(Y_i|X_i) = (1 - \pi_i)\lambda_i(1 + \pi_i\lambda_i) \tag{10}$$

If Y_i followed the ZINB distribution, then the mean and variance are given by:

$$E(Y_i|X_i) = (1 - \pi_i)\lambda_i \tag{11}$$

$$Var(Y_i|X_i) = (1 - \pi_i)\lambda_i(1 + (k + \pi_i)\lambda_i) \tag{12}$$

Where k is a parameter that indicates the over-dispersion. The ZINB model reduces itself to the ZIP one when $k \to 0$.

5 Estimatives and Results

In order to address the issue of the impact of economic inequality on repeat victimization we estimated four models: the Poisson model, the Negative Binomial Model, the Zero Inflated Poisson Model and the Zero Inflated Negative Binomial Model. However, given the inferential procedures applied to the data we opted to showed only the results[10] for the NBRM and ZINB. From Table 5 we are going to interpret the estimated elasticities and try to justify the choice of the "best" model.

First, It was verified that all estimated parameters are statistically significant at 5%, with the exception of SEX and RACE. In this sense, it appears that these variables do not influence the occurrence of repeat victimization. As a first specification test, we reject the PRM in favor of the

[10]In fact, some models are nested, and others could be tested by means of non-nested procedures

Table 5: Estimation Results

Var.	**Coef.**	**Elast.**	**Coef.**		**Elast.**
	Neg. Binomial		**Zero Inflated Neg. Binomial**		
			Zero Inflated	*Inflate*	
SIT	0.6131	0.4801	-0.2613	-1.5502	0.3944
	(21.38)	(21.38)	(-4.82)	(-18.64)	(13.20)
SEX	-0.0428^{*}	-0.0206	-0.1555^{*}	-0.2261^{*}	-0.0212
	(-1.88)	(-1.88)	(-5.02)	(-3.45)	(-1.70)
RACE	0.0407	0.0211	0.0419^{*}	0.0485^{*}	0.0093
	(2.01)	(2.01)	(1.51)	(0.83)	(0.79)
CONDD	0.9264	0.3054	0.4365	-1.5832	0.4015
	(36.04)	(36.04)	(11.80)	(-18.71)	(37.48)
AGE	0.0077	0.2602	-0.0007	-0.0145	0.2171
	(10.88)	(10.88)	(-0.69)	(-5.95)	(7.81)
ACTIVITY	0.3265	0.1884	0.1599	-0.3024	0.1785
	(14.86)	(14.86)	(5.02)	(-4.90)	(12.60)
EDUCATION	0.1037	0.4678	0.0315	-0.1777	0.5379
	(30.57)	(30.57)	(6.53)	(-16.48)	(25.51)
NUMCOMP	-0.0650	-0.3125	-0.0264	0.0646	-0.2805
	(-13.38)	(-13.38)	(-3.74)	(4.87)	(-10.59)
INCOME	0.0012	0.0433	0.0004	-0.0142	0.2487
	(9.02)	(9.02)	(3.63)	(-8.61)	(11.30)
GINI	1.9309	1.1484	2.0354	-0.3084	1.3012
	(5.51)	(5.51)	(3.96)	(-0.28)	(5.57)
POOR	0.2425	0.0994	0.1187	-0.4605	0.1419
	(3.36)	(3.36)	(1.13)	(-2.14)	(4.30)
CONSTANT	-5.1790		-2.8457	3.7975	
	(-25.67)		(-9.43)	(5.98)	
$\ln(\alpha)$	1.8347		1.3131		
α	6.2634		3.7179		
Log Likelihood = -64430.568			Log Likelihood = -63551.67		
$LR\ \ \chi^2(11) = 6980.41$			$LR\ \ \chi^2(11) = 333.85\ Prob > \chi^2 = 0.0000$		
$Prob > \chi^2 = 0.0000$			Voung test of ZINB vs. NB:		
			$z = 17.19\ \ Prob > z = 0.0000$		

NBRM as the value for the test of $H_0 = \alpha = 0$ promptly rejects this null ($LR\ \ \chi^2(11) = 6980.41$). Next, we test the ZIP model against the ZIBN model by means of a likelihood ratio test. The ZIP model is reject[11] in favor of the ZINB model. Hence, we are left with the two models presented on Table 5.

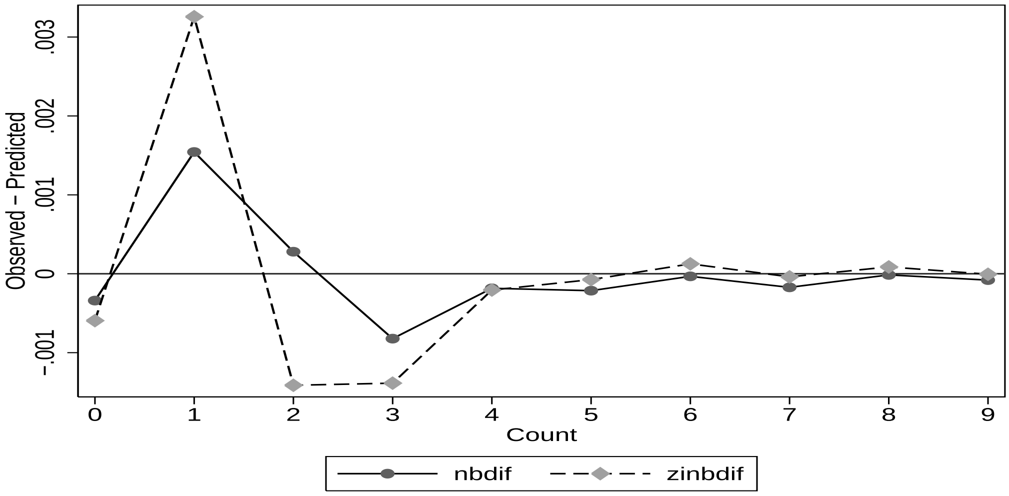
These models are non-nested, though. In order to get a first insight regarding model specification correctness, we plot the mean predicted probabilities for the NBRM and ZINB. Figure 1 plots the difference between observed probability and mean predicted probability for both model at different values of Y. The ZIP model performs much better for $Y = 0, 1, 2$ and 3, however, the ZINB outperform

[11] This results are omitted but could be obtained from the authors.

the ZIP models for higher values of Y. So, at best we reached an inconclusive result by the visual inspection of the graph.

Any more definitive result regarding the choice between these two models must be based on a non-nested test. We opted for the widespread Vuong Test. The value of this statistic appears in Table 5. The value of $z = 17.19$ flatly rejects the NBRM in favor of the ZINB. Hence, our comments on the estimated elasticities are based on our remaining "best model", i.e., the Zero Inflated Negative Binomial Model, even though the results for the NBRM are presented together. Next, we comment on the effects of the independent variables[12].

Figure 1: Mean Predicted Probabilities



Living in an urban area increases the number of victimizations, although this impact is inelastic(0.3944). Heads of the household suffer more victimizations than non-head does (0.4015), probably due to the fact that they must expose themselves more to guarantee the survival of the household, i.e., they must work more hours, perform duties related to the household, go out more often, carry more valuable objects and so on. The explanation for the effect of the variable ACTIVITY goes analogously. AGE has a positive effect as well which can be easily rationalized by noting that older people are usually richer than the younger ones.

The level of education seems to have a positive effect on the mean number of victimizations. That results, apparently at odds, reveals an interesting fact. If we assume that the probability of being a victim is function of the attractiveness (i.e., likely amount of income that could be taken) of the target and of the level of private protection (represented by technologies of protection, knowledge and behaviors) the positive elasticity seems to indicate that the prevailing net effect is the target attractiveness. This means that even though more educated people know and do protect themselves better than the less educated ones, their higher average income make them better targets to the criminal's eyes. Family size, proxied by the number of household occupants, has a negative effect on victimization. Such effect, known as guardian effect, seems reasonable as larger families could protect their members better, just by acting as guardians. Per capita family income (INCOME) has a positive

[12]In order to follow the discussion the reader must track the estimated elasticities values for the zero inflated negative binomial model. The interpretation for elasticities is straightforward: they give the percentage change on the mean expected number of victimizations given a small percentage change on the independent variable. These values are located at the last column with t-values under parenthesis, at the right.

elasticity (0.2487). So far, with few exceptions, our results confirm those obtained in the international literature, for instances, Tseloni and Pease (2003).

The level poverty has a positive, although very inelastic effect (0.1419) on repeat victimization. This suggest that poverty might not be the key issue on victimization. Note that this evidence is not at odds with a lot of evidence suggesting that criminals are more likely to be those ones with lower incomes. Note that in order to have a episode of victimization we must have a criminal, a victim and (lack of) police. We could have a situation where an increase in poverty (meaning an increase in the number of poor households) increases the number of criminals, increases the number of victimizations but, due to the "dark figure" of crime statistics, criminal reporting increases at a lower rate. We turn now to the effect of income inequality on repeat victimization.

The question regarding inequality of income and crime first appeared in a systematic approach on Ehrlich (1973). Ehrlich showed, both theoretically as well as empirically, a positive correlation between crime and income inequality. More recently Fajnzylber, Lederman, and Loayza (2000) reach the same conclusion by means of a cross-country empirical analysis. However, as far as we know there is no past attempt to measure the impact of inequality on repeat victimization. We find a strong positive effect of inequality, measured by the Gini index[13]. In fact, the estimated parameter for the GINI variable, is positive and elastic! Roughly, a 1% increase in inequality generates approximately an increase of 1.3% increase in the mean number of victimizations.

This evidence adds to the list of deleterious effects of income inequality and extends the scope of public policy. By establishing a link between inequality and victimization, governments might think more carefully about policies focused on crime prevention (victimization). The traditional view that crime must be fight with police expenditure must be compared to the returns of spending in policies targeted at decreasing income inequality. Next section concludes and offers some avenues of future research.

6 Final Considerations

The objective of the paper was to analyze the socioeconomic characteristics of individuals, also considering the income inequality indicators, and to verify how much the they can influence in the repeat victimization. Using the data from the 1988 PNAD and the IPEADATA for the same year, it was possible to create an econometric model for counting data, to apply it to the repeat victimization phenomenon and estimate it. The results of this investigation shown compatible with the international literature on the subject.

The characteristics of individuas was shown to be important to repeat victimization. Living in urban areas turns the person more vulnerable to suffer some robbery or thievery crimes. This result was expected, since great part of crimes of occurs in such areas. Moreover, being male, family head and to have some type of job also increase the person's probability to be victimized.

Years of studies, that, in accordance with economic literature, possess positive relation with the person income, also have a positive impact on repeat victimization. It is not difficult to understand this, since more time of schooling results in person having more qualifications and more opportunities, this in turn increases his/her income. Hence, such kind of individuals become possible targets to criminals. It was suggested in Meier and Miethe (1993) that the number of components of a family can help diminish the victimization risks. Considering this affirmation, the estimations confirmed this empirical evidence, showing a positive relation impact on repeat victimization.

[13] We found similar results using the Theil index. The results could be requested from the authors.

Gini index has a positive relation with repeat victimization and this effect is the highest among all other variables. However, important sources of variations were not employed in the present study. An important omission is discussion about the separate effects of unobserved heterogeneity and event dependence (see, Tseloni and Pease (2004)).

In face of all that, we can be conclude that repeat victimization occurs mainly due to socioeconomic conditions of the individual and the region. Moreover, the possibility of being victimized is bigger for those who had already been victims sin the past. Due to this, it is important to develop a notion that the repeat victimization reduction is related to the prevention and the reduction of crimes. The results of this work had become evident that such programs cannot be based only on momentary questions, such as to increasing the police contingent in the streets, improving its armament and/or constructing new penitentiaries, but also policy makers must pat attention to questions that have long lasting effects, through policies that seek economic and social improvements in the region. In fact, in accordance with the project of Brazil (2003), in order to give an end to criminality and victimization, is necessary to offer security for all the population. Due to this, the government must be committed to create polices that fight against the main variable that lead to the increase of the violence. Such variables are mainly centered in the economic, social and of public security spheres.

Some important improvements in the paper remain to be done, though. First, we believe that multilevel model should be an important neglected issue. Given the natural clustering of observation inside the household, and the evident behavioral implications for crime prevention, a count model that incorporates a multilevel (in our case, two levels) is of paramount importance. This is especially relevant because multilevel count data is not a novelty in the criminological community, for instances, see Tseloni and Pease (2003) and Tseloni and Pease (2004). Second, as already mentioned, the issue of unobserved heterogeneity ***versus*** state dependence was far from complete in our paper. This represents a challenge, however, because we do not have longitudinal observations. Different approaches must be work out. Third, cross-effects as well as polynomials effects must be included, specially involving the variables AGE, EDUCATION, INCOME and GINI. However, this is not a trivial task as these effects are not as easily obtained as the case of linear models. Good references about this very important, but almost entirely neglected, issue are Norton, Wang, and Ai (2004) and Ai and Norton (2003).

References

Ai, C., and E. C. Norton (2003): "Interaction terms in logit and probit models," *Economics Letters*, (80), 123–129.

Araújo, A., and P. Fajnzylber (2001): "O que causa a criminalidade violenta no Brasil? uma análise a partir domodelo econômico do crime: 1981 a 1996," Texto de discussão 162, CEDEPLAR/UFMG.

Barslund, M., J. Rand, F. Tarp, and J. Chiconela (2005): "Understanding Victimization: The Case of Mozambique," Discussion Papers 5, 17, University of Copenhagen.

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

Bourguignon, F., F. Sanchez, and J. Nunez (2003): "A structural model of crime and inequality in Colombia," *European Economic Association*.

Brazil (2003): "Projeto Segurança Pública para o Brasil," Discussion paper, Fundação Djalma Guimarães.

Cameron, A. C., and P. K. Trivedi (2005): *Microeconometrics: methods and applications*. Cambrigde University Press.

Ehrlich, I. (1973): "Participation in illegitimate activities: a theoretical and empirical investigation," *Journal of Political Economy*, 3(81).

Fajnzylber, P., and et al. (1998): "Determinants of Crime rates in Latin American and the world, an empirical Assessment," Discussion paper, The Word Bank.

Fajnzylber, P., D. Lederman, and N. Loayza (2000): "What causes violent crime?," *European Economic Review*.

Farrell, G., and K. Pease (2001): *Repeat victimization*. Criminal Justice Press.

Gaviria, A., and C. Pagés (1999): "Patterns of crime victimization in Latin American," Working Paper 408, Inter-American Development Bank.

Gomes, F., and L. S. Paz (2007): "The Determinants of Criminal Victimization in São Paulo state-Brazil," Discussion Paper WPE 25-2007, IBMEC.

Kleemans, E. R. (2001): "Repeat burglary victimization: results of empirical research in the Netherlands," in *Repeat Victimization*, ed. by G. Farrell, and K. Pease, vol. 12. Criminal Justice Press.

Latinobarometer (1999): "Data set," www.lationobarometro.org.

Laycock, G. (2001): "Hypothesis-based research: the repeat victimization story," *Criminal Justice*, 1(1), 59–82.

Levitt, S. D. (1999): "The changing relationship between income and crime victimization," *Economic Policy Review*, (september).

MacDonald, Z. (2001): "Revisiting the dark figure: a microeconometric analysis of the under-reportingof property crime and its implications," *British Journal of Criminology*, 41.

——— (2002): "Official crime statistics: their use and interpretation," *The Economic Journal*, 112.

Meier, R. F., and T. D. Miethe (1993): "Understanding theories of criminal victimization," Discussion paper, University of Chicago.

Norton, E. C., H. Wang, and C. Ai (2004): "Computing interaction effects and standard errors in logit and probit models," *The Stata Journal*, 4(2).

Pease, K. (1998): "Repeat victimization: taking stock," Series Paper 90, London: Home Offica.

Sparks, R. F. (1981): "Multiple victimization: evidence, theory, and future research," *Journal of Criminal Law and Criminology*, 72(2).

Tella, R. D., S. Galiani, and E. Schargrodsky (2002): "Crime victimization and income distribution," Discussion paper, Inter-American Development Bank.

Thacher, D. (2004): "rich get richer and the poor get robbed: inequality in U. S. criminal victimization,1974-2000," *Journal of Quantitative Criminology*, 20(2).

Tseloni, A., and K. Pease (2003): "Repeat personal victimization. "Boosts" or "Flags"?," *British Journal of Criminology*, (43).

——— (2004): "Repeat personal victimization: random effects, event dependence and unexplained heterogeneity," *British Journal of Criminology*, 44, 931–945.

Ybarra, L. M. R., and S. L. Lohr (2002): "Estimates of repeat victimization using the Nacional Crime VictimizationSurvey," *Journal of Quantitative Criminology*, 18(1).