An estimate of the underreporting of violent crimes against property applying stochastic frontier analysis to the state of Minas Gerais, Brazil

Estimativa do subregistro de crimes violentos contra a propriedade aplicando análise de fronteira estocástica para o estado de Minas Gerais, Brasil

Gustavo Carvalho Moreira [1]
Ana Lucia Kassouf [2]
Marcelo Justus [3]

[1] Universidade Federal de São João del Rei
[2] Universidade de São Paulo
[3] Universidade de Campinas

Abstract

The aim of this study was to determine the level of underreporting of violent crimes against property in the state of Minas Gerais, Brazil, using stochastic frontier analysis. Failure to report a crime to competent authorities has negative consequences for the effectiveness of public safety policies, because policy decisions regarding criminal matters are based on official crime statistics, which in turn are biased by underreporting. An awareness of the magnitude of underreporting should help policy makers to design more appropriate crime prevention strategies. The database used for the study defines armed robbery, robbery, and theft as violent crimes against property. The main results of the study show that, from 2004 to 2011, 32.7% of all violent crimes against property in the state were not reported. A robustness test was carried out on the estimates by employing the same technique to calculate the level of underreporting for the homicide rate, which is the least unreported crime.

Resumo

O objetivo desse estudo foi determinar o nível de subregistro de crimes contra a propriedade no estado de Minas Gerais, Brasil, utilizando a técnica de análise de fronteira estocástica. Não comunicar um crime às autoridades competentes gera consequências negativas para a efetividade das políticas de segurança pública, uma vez que a tomada de decisão é realizada tendo como base as estatísticas oficiais de crimes, que serão viesadas pelo subregistro. Neste sentido, tomar conhecimento sobre o nível de subregistro é fundamental para aprimorar as estratégias de prevenção ao crime. Foram utilizados dados referentes aos crimes de roubo à mão armada, roubo e furto. Os resultados indicaram que o subregistro médio dos crimes analisados foi de 32,7% no período entre 2004 e 2011. Um teste de robustez foi realizado nas estimativas, empregando a mesma técnica para calcular o nível de subnotificação para homicídios, que é o crime com o menor nível de subregistro.

Keywords

crime prevention; economy of crime; public safety.

JEL Codes C35, K14, I31.

Palavras-chave

estratégias de prevenção ao crime; economia do crime; segurança pública.

Códigos JEL *C35, K14, I31.*

1 Introduction

A crucial problem that arises when analysing criminality is the fact that criminal statistics are subject to measurement errors, mainly due to crime underreporting. For example, a crime will only become part of the crime statistics when the victim reports it to the competent authorities. Unreported crime causes measurement errors that skew the perception of the true crime rate, which is a serious problem when allocating resources to combating and preventing crime.

Given the problem of statistical bias caused by crime underreporting, some studies estimate the true crime rate using victimization surveys. Victimization surveys emerged in the United States in the 1960s with the objective of measuring and analysing crimes suffered by the population that, for some reason, were not reported. In Brazil, the first national victimization survey was carried out by the Brazilian Institute of Geography and Statistics (IBGE) in 1988. The most recent surveys were conducted by the IBGE in 2009 and by the National Public Security Secretariat in 2012. As victimization surveys have become essential to the determination of true crime rates, the results from further studies that increase the understanding of the victimization phenomenon, such as those conducted by Justus and Kassouf (2008a), Madalozzo and Furtado (2011), Peixoto *et al.* (2012), Justus and Kassouf (2013), and Sant'Anna *et al.* (2016), should lead to more efficient and productive public policies to address crime.

Although victimization surveys are very rich in information, Justus and Kassouf (2008a) point out that dependence on their results has drawbacks. Among these disadvantages, we can highlight the fact that no victimization surveys in Brazil are carried out with periodicity, that the surveys that are performed do not use a sampling method defined by a single entity, which makes comparisons impossible, and that the surveys' high costs severely limit their execution.

An alternative method to measure underreporting is based on the application of stochastic frontier analysis; however, the data derived using this methodology do not replace the wealth of information obtained from victimization surveys. Chaudhuri *et al.* (2015) employ stochastic frontier analysis to estimate an economic model that explains variations in crime rates. Their analysis allows one to isolate the error term in a deterministic component under a certain set of hypotheses defined a priori. The deterministic

term, called technical inefficiency in the case of a production function, represents the distance between the observed crime rate and the frontier. According to the authors, this distance is classified as the share of crimes that are not reported. Their study of India, examining a crime index comprised of murder, attempted murder, manslaughter, rape, kidnapping and abduction, armed robbery, and robbery, found an underreporting average of 27%.

Although less detailed than victimization surveys, stochastic frontier analysis can complement these surveys while providing more standardized and more frequent estimates of underreporting. The objective of this study was to follow Chaudhuri *et al.* (2015) and apply stochastic frontier analysis to measure crime underreporting in the state of Minas Gerais, Brazil, from 2004 to 2011. The crimes analysed in our study were crimes against property resulting from robbery, armed robbery, and theft. The results obtained indicate that, over the study period, on average, 32.7% of these crimes in the state were not reported, a value close to that verified by previous victimization surveys.

This study is divided into four more sections. Section 2 presents the theoretical justifications and the likely effects of crime underreporting. Section 3 contains the methodological framework and empirical models. Section 4 presents the underreporting estimates and a discussion of these results. Section 5 offers concluding remarks.
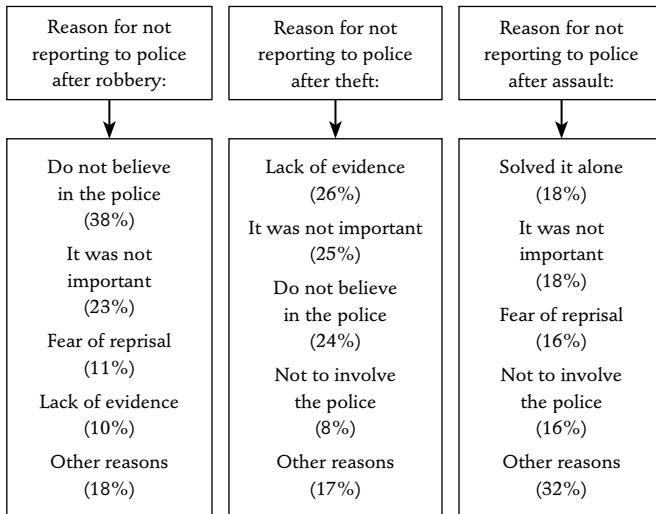
2 Theoretical justifications and the likely effects of crime underreporting

There are two main hypothesized justifications for not reporting a crime. First, rational choice emphasizes that victims of crime consider the benefits and costs of reporting the crime. The benefits are related to police protection, the possibility of the criminal being detained, and a reduction of the risk of future victimization. The costs, in turn, are the inconveniences associated with legal proceedings, the fear of revenge and retaliation, and the complaint's opportunity cost. The victim will report the crime only if the net result of this cost–benefit analysis is positive. Empirical evidence of the relationship between costs and benefits of crime reporting can be found in the studies by Soares (2004a), Justus and Kassouf (2008a), and Madalozzo and Furtado (2011).

The second reason for not reporting a crime is that the victim may not trust the authorities to which the crime should be reported. This issue is brought up in other crime underreporting studies, among which are those by Silver and Miller (2004), Soares (2004b), Goudriaan *et al.* (2005), and Warner (2007).

Figure 1 presents the reasons that individuals give for not seeking Brazilian police assistance after being a victim of robbery, theft, or assault. It is interesting to note that the victim's reasons for not reporting a crime against property to the proper authorities are prioritized differently depending on the type of crime.

Figure 1 **Reasons given to avoid involving the police after victimization, by crime type**

Reason for not reporting to police after robbery:	Reason for not reporting to police after theft:	Reason for not reporting to police after assault:
Do not believe in the police (38%)	Lack of evidence (26%)	Solved it alone (18%)
It was not important (23%)	It was not important (25%)	It was not important (18%)
Fear of reprisal (11%)	Do not believe in the police (24%)	Fear of reprisal (16%)
Lack of evidence (10%)	Not to involve the police (8%)	Not to involve the police (16%)
Other reasons (18%)	Other reasons (17%)	Other reasons (32%)

Source: Brazilian Institute of Geography and Statistics (IBGE) (2009).

Crime underreporting has three main likely effects: it causes statistical errors, it leads to the misallocation of resources, and it alters the criminal's cost–benefit analysis. Justus and Kassouf (2008a) indicate that this phenomenon precludes a precise interpretation of crime statistics. Figure 2, for example, shows that the rate of violent crimes against property in Minas Gerais increased during the period from 2001 to 2007. However, this trend could stem from two sources other than an increase in crime: the police force's efficiency in combating these crimes improved and more arrests were made or the number of reports of these crimes increased.

Figure 2 **Number of violent crimes against property in Minas Gerais per 100 thousand inhabitants for the years 2001 to 2011**
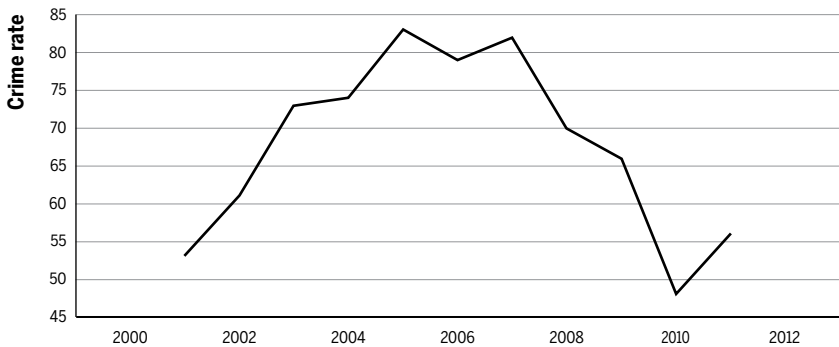
Source: João Pinheiro Foundation (2015).

Underreporting affects the allocation of resources to the improvement of public safety and may cause the allocation to be suboptimal. In general, this allocation is made on the basis of reported crime indicators, which can have dubious interpretative value, and underreporting would cause a distortion in the amount of resources allocated to the prevention and prosecution of crime. To illustrate, consider the hypothetical municipal units shown in Table 1. Apparently, municipality 1 would have the least need for public safety resources because it had the lowest reported crime rate (90). However, the high rate of underreporting in this municipality (50%) makes the true crime rate the highest among the four municipalities. For municipality 4, the resources earmarked for the prevention of crime are most likely to be the highest, yet, in the face of zero underreporting, there is less need for public safety resources in this municipality than in the other municipalities.

Table 1 **Reported crime rate and underreporting for hypothetical municipalities**

Municipality	Crime rate	Underreporting (% of crime rate)	Corrected crime rate
1	90	50	135
2	100	30	130
3	110	10	121
4	120	0	120

Finally, Becker's (1968) economic model of crime emphasizes agents' economic rationality. According to this model, all individuals are potential offenders, and the decision to commit a crime is made by assessing the relationship between the return on illegal activity, the opportunity cost, and the severity of the penalties. If individuals do not report crime after victimization, no police inquiry is generated and therefore there is less likelihood of the criminal being captured. Under this assumption, underreporting would contribute to increasing the crime's potential return, if being prosecuted is considered to affect the return negatively. This argument is also discussed by Justus and Scorzafave (2014).

3 **Method**

3.1 **Stochastic frontier analysis**

Stochastic frontier analysis (SFA) is also relevant to the study of other phenomena. With respect to the social economy, SFA is frequently used to analyse the labour market to determine how much the individual wage distances itself from labour's marginal productivity (Groot and Ooterbeek, 1994; Polachek and Robst, 1998; Prieto *et al*., 2005; Adamchik and King, 2007; Sánchez, 2011). Kumbhakar *et al*. (2015) provide several other applications of SFA to analyse efficiency in their aptly titled tract 'A Practitioner's Guide to Stochastic Frontier Analysis Using STATA'.

In microeconomic theory, a production function is defined as the maximum amount of products that can be generated given a set of inputs and available technology. Inefficiency may arise for many reasons, among them a lack of motivation, asymmetric information, ineffective monitoring, unoptimized market structures, and management problems. According to Kumbhakar and Lovell (2000), the seminal works by Debreu (1951) and Shepard (1953) seek to estimate a distance function theoretically to capture a unit's degree of inefficiency. As the methodology advanced and new non-parametric linear programming techniques were developed, more robust inefficiency estimation methods appeared, such as data envelopment analysis (DEA) and deterministic frontier analysis (DFA). Coelli (1995) points out that both types of analyses can be criticized due to the fact that the measurement of efficiency cannot be controlled by the ran-

dom components that affect it. In the case of DFA, for example, this occurs because the error term (u_i) in the estimation of a production function described by

$$y_i = \beta' x_i - u_i \tag{1}$$

is classified as being a component of technical inefficiency, whereas it is also affected by the stochastic term (which captures measurement errors, omitted variables, and other factors). In eq. (1), y_i indicates the production of unit i, x_i is the input vector used by unit i, and β is the parameters of technology.

To circumvent the fact that the technical inefficiency term (u_i) is also affected by random noise, stochastic frontier analysis adds a random component, v_i, to the error term. In this way, eq. (1) becomes:

$$y_i = \beta' x_i - \varepsilon_i \tag{2}$$

with $\varepsilon_i = v_i - u_i$ and v_i and u_i being independently distributed between them and between the regressors. From this change, the first component of the error term, v_i, is assumed to be random, with mean 0 and variance σ_v^2, with a density function given by:

$$f(v) = \frac{1}{\sqrt{2\pi\sigma_v^2}} . exp\left\{ -\frac{v^2}{2\sigma_v^2} \right\} \tag{3}$$

In turn, the second component of the error term u_i is a non-negative random variable of which the distribution behaviour is defined a priori and captures the effect of technical inefficiency. The assumptions for the behaviour of u_i take several forms, such as truncated normal with mean μ and variance σ_v^2, exponential, truncated in 0, and gamma (Coelli et al., 2005). In the case of the truncated normal form, the distribution function of u is defined by

$$f(u) = \frac{1}{\sqrt{2\pi\sigma_u^2 \odot \left(-\dfrac{\mu}{\sigma_u^2} \right)}} . exp\left\{ -\frac{(u-\mu^2}{2\sigma_u^2} \right\} \tag{3.1}$$

with μ being truncated below zero and $\odot$ (.) being a cumulative normal distribution function. Under these assumptions, eq. (2) is then obtained by maximum likelihood, and it is necessary to estimate three additional parameters, namely μ, σ_u^2, and σ_v^2. The log-log function to be estimated for I producers can be described by eq. (4):

$$lnL\left(y|\beta,\mu,\sigma\right) = constant - Iln\sigma - Iln \odot\left(-\frac{\mu}{\sigma_u^2}\right) +$$

$$+ \sum_i ln \odot\left(\frac{\mu}{\sigma\lambda} - \frac{\varepsilon_i\lambda}{\sigma}\right) - \frac{1}{2}\sum_i ln\left(\frac{\varepsilon_i + \mu}{\sigma}\right)^2 \tag{4}$$

where $\sigma = \left(\sigma_u^2 + \sigma_v^2\right)^{1/2}$, $\lambda = \dfrac{\sigma_u}{\sigma_v}$, and $\varepsilon_i = v_i - u_i$. Further details about this

and other forms of estimation, depending on the assumed distribution, can be found in Kumbhakar and Lovell's (2000) book.

Coelli *et al.* (2005) point out that the most common way of predicting u_i (the term of technical inefficiency), after the estimation of eq. (4), is given by eq. (5), which represents the ratio of the observed output (y_i) to that obtained by stochastic frontier analysis:

$$inefficiency_i = \frac{y_i}{\exp\left(x_i'\beta + v_i\right)} = \frac{\exp\left(x_i'\beta + v_i - u_i\right)}{\exp\left(x_i'\beta + v_i\right)} = \exp\left(-\widehat{u_i}\right) \tag{5}$$

where $\widehat{u_i} = E(u_i | u_i + v_i)$.

Similarly, the above explanation can be extended to a panel data analysis using a general equation, such as eq. (6):

$$y_{it} = \alpha + \beta' x_{it} + v_{it} - u_{it} \tag{6}$$

with the term u_{it} being able to vary over time or not. If the term varies over time according to a function $f(t)$, we have:

$$u_{iT} = u_i f(t) \tag{7}$$

in which, according to Battese and Coelli (1992), $f(t)$ is given by eq. (8):

$$f(t) = \exp\left(-\gamma(t - T)\right) \tag{8}$$

with γ being a parameter to be estimated and t representing the unit of time within a set of times T. If $\gamma > 0$, inefficiency decreases over time. A description of the maximum likelihood function, in the case that the term of inefficiency varies over time, is provided by Kumbhakar and Lovell (2000, p. 111). Their hypothesis is that inefficiency (*inef*) varying over time can be predicted using an equation such as eq. (9):

$$\exp\left(-u_{it}\right) = inef_{it} = inef_i \exp\left(-\eta\left(t - T\right)\right) \tag{9}$$

It is important to note that the application of SFA has some limitations, like any statistical technique. As noted by Bezat (2009), the assumed functional form is defined a priori, the choice of an incorrect specification can influence the empirical results, the estimation does not allow one to obtain inferences for small samples, and u_i is sensitive to the type of distribution assumed for its behaviour.

3.2 Data

The data used in this study came from the João Pinheiro Foundation. The foundation's database contains information on health, education, public safety, municipal finance, environment and housing, sports and tourism, income and employment, social assistance, and culture for all of Minas Gerais's 853 municipalities from 2001 to 2011. The definition of periodicity, as well as the choice of the region to study, was motivated by the data availability at the time of this study. The exercise performed here can clearly be applied to other units of analysis, subject to the data availability.

3.3 Selected variables

The regression model defined by eq. (6) was constructed to estimate violent property crime underreporting through the use of a stochastic frontier model, with the rate of violent crimes against property acting as the dependent variable. Violent crimes against property are defined by the study's João Pinheiro Foundation database as follows:

The ratio between the number of recorded occurrences of violent crimes against property and the population of the municipality, multiplied by 100,000.

According to the database and the Brazilian Penal Code, theft, robbery, and armed robbery are considered to be violent crimes against property.

The selection of explanatory variables for this model was based on Becker's (1968) theoretical framework and other empirical studies on the subject. Table 2 briefly describes these variables as well as the expected sign for their coefficients. The variables' descriptive statistics can be found in Table A1 of the Appendix.

Table 2 **Description of the variables used in the empirical model and the expected sign of the estimated coefficients**

	Description	E.S.
police	Number of military police officers in the municipality per 100 thousand inhabitants	–
income	Value of the total income of formal employees in the month of December divided by the number of employees	–
GDPpercapita	Gross domestic product per capita	?
young	Number of individuals residing in the municipality between 15 and 29 years of age per 100 thousand inhabitants	+
urban	Percentage of the population of the municipality living in an urban area	+
vulnerability	Value of transfers from the Bolsa Família welfare programme in the year per 100 thousand inhabitants	+
Wcrime	Spatially lagged violent crime against property rate per 100 thousand inhabitants	+
crime(-1)	Lagged dependent variable	+

Note: E.S.: expected sign; all monetary values were deflated by the IGP-DI at 2011 prices.

The economic model of criminal rationality developed by Becker (1968) assumes that an individual will decide whether or not to commit a crime by evaluating the activity's return subject to its cost. The severity of the penalties and the probability of punishment are considered to be costs and are controlled in the model by the variable *police*, which indicates the number of military police officers in the municipality per 100 thousand inhabitants. The empirical studies by Araújo Junior and Fajnzylber (2001), Loureiro and Carvalho Junior (2007), Loureiro (2009), and Teixeira (2011) also control for this aspect using other proxies, such as public safety spending.

The opportunity cost of illegal activity is another item considered to be part of the cost of criminal activity and is controlled in the model by the variable *income* and the variable *vulnerability*. *Income* represents the value of the total income of formal-sector employees in December of the selected year, divided by the number of employees. The variable *vulnerability* refers to the value of transfers from Bolsa Familia, a public assistance programme, per 100 thousand inhabitants. In the case of the *income* variable, it is hypothesized that higher wages in the formal sector will reduce the incentives for those employed in this sector to engage in illegal activities against property, although they could also be considered as an incentive to engage in criminal activity because the potential victims have become more attractive (Justus and Kassouf, 2008b). For the *vulnerability* variable, it is expected that greater socioeconomic vulnerability will create an environment that is more conducive to crime. Using other proxies, the latter relation is empirically verified by Carvalho and Lavor (2008), Resende and Andrade (2011), and Shikida and Oliveira (2012), among others.

The return from criminal activity is controlled by the *GDPpercapita* variable, a proxy for the municipality's wealth. The expected sign for the relationship between a municipality's wealth and the return from criminal activity is ambiguous. On the one hand, higher income implies a population with greater purchasing power that eventually allows access to a basket of consumption goods and services composed of safety and defence items that inhibit criminal activity. On the other hand, higher income implies potentially more attractive victims (see Cohen *et al.*, 1981), which may increase the return from criminal activity.

The *young* variable represents the number of residents in the municipality aged 15–29 per 100,000 inhabitants. The variable captures the influence of this age group on the crime rate. The inclusion is motivated by the fact that 55% of homicides during a conflict situation (assaults, maltreatment, or abandonment) in 2013 involved people between 15 and 29 years of age (International Classification of Diseases X85–Y09).

The variable *urban* is another demographic control. It represents the percentage of the population living in an urban area. According to Teixeira (2011), where an individual lives affects the chance that he or she will become involved in criminal activity, hypothesizing that urban settings diminish Becker's punishment effect (Becker, 1968). Teixeira (2011) reasons that the chance of being caught is perceived by criminals as unlikely in

urban areas and that the opportunity for criminals to encourage and train future criminals is more likely than in areas with lower population density.

In recognition of the empirical evidence that criminal activities tend to agglomerate, the variable *Wcrime* is used to control for spatial effects on the crime rate. The methodology used to construct this variable is discussed in the next subsection. The empirical studies by Peixoto (2003), Almeida *et al.* (2005), and Justus and Santos Filho (2011) verify the spatial effect but only in regard to homicides. Also noteworthy is the work by Tolentino and Diniz (2014) on the spatial dependence of drug trafficking in the capital of Minas Gerais, Brazil: Belo Horizonte.

A lagged dependent variable, *crime(-1)*, was included to model the learning-by-doing phenomenon's effect on crime rates. According to Justus (2009), estimates for Brazil indicate that half of the crimes committed in one period are transferred to the next. The studies by Araújo Junior and Fajnzylber (2001), Kume (2004), and Teixeira (2011) also control for this aspect.

3.4 **Empirical modelling**

The empirical modelling began by estimating fixed-effect (FE) and random-effect (RE) baseline models using the maximum-likelihood estimation technique. Then, two more robust models were estimated to address potential spatial dependence and endogeneity problems. Mathematically, the estimated FE model is described by eq. (10):

$$ln y_{it} = \beta_{0t} + \sum_n \beta_n ln x_{nit} + v_{it} - u_{it} = \beta_{it} + \sum_n \beta_n ln x_{nit} + v_{it} \tag{10}$$

in which y_{it} represents the violent crime rate against property in municipality i for period t (*crime*), x_{nit} are the n control variables presented in Table 2, for each municipality i and period t, excluding the *crime(-1)* and *Wcrime* variables, and u_{it} is the underreporting term ($u_i \sim iidN^+(\mu, \sigma_u^2)$). β_0 is the intercept, common to all municipalities in each period t, and $\beta_{it} = \beta_{0t} - u_{it}$ is the intercept for municipality i in period t.

It should be noted that, for u_i, normal truncated behaviour of mean μ and variance $\sigma_u^2 (u_i \sim iidN^+(\mu, \sigma_u^2))$ is assumed. This term promotes flexibility in the distribution of underreporting to be estimated.

Similarly, the RE model is given by eq. (11):

$$lny_{it} = \left[\beta_{0t} - E\left(u_{it}\right) \right] + \sum_{n} \beta_n lnx_{nit} + v_{it} - \left[u_{it} - E\left(u_{it}\right) \right] =$$

$$= \beta_{0t}^{*} + \sum_{n} \beta_n lnx_{nit} + v_{it} - u_{it}^{*} \tag{11}$$

To permit the level of underreporting to vary over time in either the FE or the RE model, $f(t)$ was considered to be a function of underreporting behaviour over time, as described in equations (7) and (8).

Comparing the consistency of the estimators obtained in the FE and RE models (Table 3), Hausman's (1978) test indicates that the FE model is preferable to the RE model. According to Justus (2009), it is actually more plausible to use fixed-effect models specifically in the case of crime analysis, because the unobservable municipal characteristics are potentially correlated with the model's exogenous variables. It is important to note that, in addition to the models mentioned above, two additional specifications were estimated and tested, DEF and END. The DEF model is the fixed-effect model (FE) with a spatially lagged crime control. The END model is the DEF model with an addition that addresses endogeneity. The results from the estimation of all four models are presented in Table 3.

Table 3 **Results from the empirical models**

	RE	FE	DEF	END
police	0.0088	0.0088	*0.0016	***-0.0004
	(0.0159)	(0.0159)	(0.0010)	(0.0000)
income	***-0.0120	***-0.0120	***0.0003	***0.0300
	(0.0019)	(0.0019)	(0.0001)	(0.0060)
GDPpercapita	**-0.0002	**-0.0002	-0.0001	*0.0267
	(0.0000)	(0.0000)	(0.0001)	(0.0143)
young	*-0.0007	*-0.0007	**0.0006	***0.0001
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
urban	0.1012	0.1012	***3.0476	***0.0030
	(0.0809)	(0.0809)	(0.6452)	(0.0003)
vulnerability	**-0.0206	**-0.0206	*0.0520	***0.00002
	(0.0094)	(0.0094)	(0.0359)	(0.0000)
Wcrime			***0.1666	***0.3389
			(0.0178)	(0.0120)
crime(-1)				***0.0058
				(0.0000)

(continues on next page)

Table 3 **(continued)**

|	RE	FE	DEF	END
N	9,097	9,097	9,097	6,616

Notes: Standard errors, estimated by municipality clusters, are described in parentheses; ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Source: Authors.

After the estimation using the FE model, the next step was to employ the DEF model to control for the crime rate's spatial effect. The studies by Peixoto (2003), Almeida *et al*. (2005), and Justus and Santos Filho (2011) find that a region with a high crime rate tends to influence the crime rates in neighbouring regions; that is, the crime rate is not spatially independent. In addition, according to Almeida (2013), the non-inclusion of the spatial gap violates the basic hypotheses of the BLUE estimators, implying estimator bias and model inconsistency. Figure 3 shows evidence that the crime rate in Minas Gerais is spatially dependent: municipalities with high crime rates are surrounded by others with the same crime characteristics.

Figure 3 **Rate of violent crimes against property in Minas Gerais per 100 thousand inhabitants in 2011**
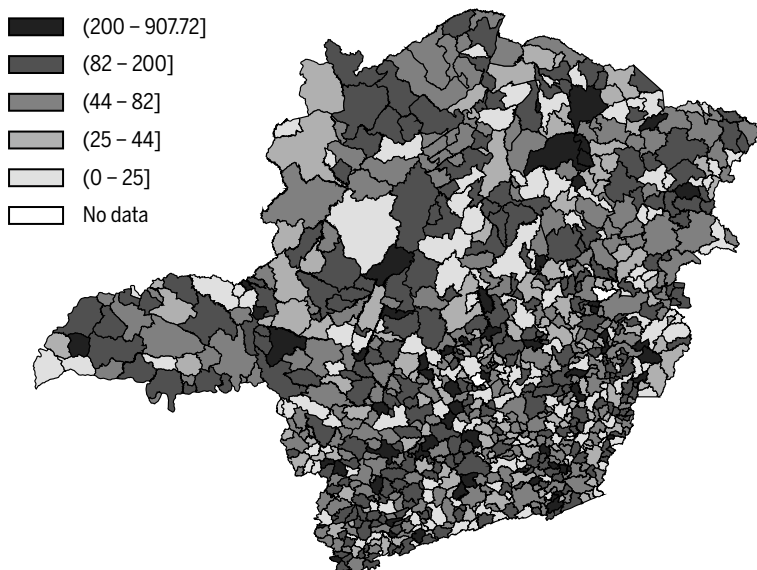
Source: João Pinheiro Foundation (2015).

Despite the evidence and intuition regarding the crime rate's spatial dependence, a test was conducted to verify empirically the existence of this dependency in Minas Gerais's municipalities. In this case, the Moran I was calculated according to eq. (12):

$$I = \frac{n}{S_0}\left(\frac{z'Wz}{z'z}\right)$$

(12)

where n represents the number of regions (the state's 853 municipalities), z represents the vector of values of the variable of interest, W is the spatial weighting matrix, and S_0 is the sum of all the elements of the matrix of spatial weights $W(S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n}w_{i,j})$. The value z_I, used to test for the presence of spatial correlation, was computed by eq. (12.1):

$$z_I = \frac{I - E(I)}{\sqrt{V(I)}}$$

(12.1)

in which $E(I)$ and $V(I)$ were defined by eq. (12.2) and eq. (12.3), respectively:

$$E(I) = {-1}/{(n-1)}$$

(12.2)

$$V(I) = E(I^2) - E(I)^2$$

(12.3)

An 853 × 853 square was adopted as the spatial weighting matrix defined according to eq. (12.4):

$$W = \begin{bmatrix} 0 & w_{kj}(d_{k,j}) & \cdots & w_{kl}(d_{k,l}) \\ w_{jk}(d_{j,k}) & 0 & \cdots & w_{jt}(d_{j,t}) \\ \cdots & \cdots & \cdots & \cdots \\ w_{lk}(d_{l,k}) & w_{lj}(d_{l,j}) & \cdots & 0 \end{bmatrix}$$

(12.4)

in which each element represents the inverse Euclidean distance to the normalized square, which was calculated using the municipalities' latitude

and longitude ($i, j, \ldots, t$), obtained directly from the IBGE. The distance between municipality i and municipality j is defined by the system represented in eq. (12.5):

$$w_{ij}^* = \begin{cases} \dfrac{1}{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)\dagger}, if \ i \neq j \\ 0, if \ i = j \end{cases}$$

(12.5)

where x_i, x_j, y_i, y_j are the centroid coordinates of units i and j. To obtain the components of the matrix W, finally, each of the elements was normalized according to eq. (12.6):

$$w_{ij} = \frac{w_{ij}^*}{\sum_j w_{ij}^*}$$

(12.6)

According to Tsyzler (2006), the inverse Euclidean distance is most often used in spatial econometric analyses and is of the type 'everyone influences but those who are closer influence more'. Figure 4 shows the crime rate's dispersion for the years 2001 and 2011 in relation to their respective spatial lag. The cluster of observations indicates a spatial correlation between the municipalities' crime rates.

In addition, the robust Lagrange multiplier (LM) test was performed to detect spatial autocorrelation (Table 4) and to determine whether the model to be specified should contain a spatial lag in the dependent variable, in the error term, or in both (see Justus and Santos Filho, 2011, pp. 139–140). At the significance level of 1%, the hypothesis of the absence of spatial autocorrelation was rejected only in the dependent variable: violent crimes against property.

Table 4 **Robust Lagrange multiplier (LM) test for spatial dependence**

Variable	LM test	P-value
crime	6,1170	0,0000
error term	0,7010	0,4024
crime and error term	0,7010	0,7043

Note: Null hypothesis: absence of spatial autocorrelation.

Source: Authors.

From the results shown in Table 4, we have a spatial autoregressive (SAR) stochastic frontier model that is defined by eq. (13) and satisfies the assumptions presented in eq. (13.1):

$$lny_{it} = \rho \mathbf{W} y_{it} + \beta_{0t} + \sum_n \beta_n lnx_{nit} + v_{it} - u_{it} \tag{13}$$

$$E(v_i t) = 0; E(v_i v_{t'}) = \sigma^2 \mathbf{I}_n \tag{13.1}$$

in which $Wy_{it} = Wcrime_{it}$ represents the crime rate spatially lagged by the distance matrix W and ρ is the spatial correlation parameter to be estimated ($1 \le \rho \le -1$). The term $\sum_n \beta_n lnx_{nit}$ contains all the variables described in Table 2, with the exception of the *crime(-1)* variable. The estimates can be found in Table 3 in the DEF column.

Finally, the problem of endogeneity was addressed, as it affects both traditional econometric models and stochastic frontier estimates. There is potential endogeneity for four of the explanatory variables listed in Table 2: i) correlation between the lagged dependent variable *crime(-1)* and the error term; ii) correlation between the spatially lagged dependent variable (which is used as explanatory) and the error term; iii) simultaneity between the *police* and the dependent variable; and iv) simultaneity between the *income* variable and the dependent variable. To overcome this problem, we followed the procedure developed by Tran and Tsionas (2013), a procedure that asymptotically resembles the two-stage residual inclusion method proposed by Terza *et al.* (2008).

Based on Tran and Tsionas (2013), we considered a general stochastic frontier model with an endogenous regressor according to eq. (14) and eq. (14.1):

$$lny_{it} = \rho \mathbf{W} y_{it} + \beta_{0t} + \sum_n \beta_n lnx_{nit} + \sum_n \beta_n lnz_{nit} + v_{it} - u_{it} \tag{14}$$

$$lnx_{it} = lnz_{2,it} + \varepsilon_{it} \tag{14.1}$$

in which $\sum_n x_{nit}$ is a set of endogenous regressors, $\sum_n z_{nit}$ is a set of exogenous regressors, and $z_{2,it}$ is assumed to be a strictly exogenous instrument. The procedure adopted was to apply the two-stage residue inclusion method adopted by Terza *et al.* (2008): first, predict ε_{it} in eq. (14.1), then replace the endogenous term of eq. (14), x'_{it}, with the estimated ε_{it} from (14.1).

Following Arellano and Bond's (1991) procedure for the correction of an endogenous variable, the variable itself was lagged over two and three periods as an instrument, thereby satisfying the condition of being correlated with the endogenous variable and not being correlated with the error term. After adopting this procedure, a new model was estimated considering both spatial dependence and endogeneity, the results of which can be found in the END model column of Table 3. The END model was determined to be the most inclusive model.

After constructing the empirical models, the predicted values of $\hat{u}_{it}$, as described by eq. (9), were estimated to capture the rate of crime under-reporting.

4 Results and discussion

Table 5 presents descriptive statistics of crime underreporting as estimated by the fixed-effect model (FE), the FE model with spatially lagged crime control (DEF), and the FE model that is both spatially lagged and controlled for endogeneity (END). Estimates were generated for only 827 of the state's 853 municipalities, as there was no information in the João Pinheiro Foundation data set for the 26 remaining municipalities.

The END model was selected as this study's main model. That model controls for the problems of spatial autocorrelation and endogeneity and was found to be the most complete. It was determined that, if spatial autocorrelation and endogeneity problems were disregarded, the model would contain biases and inconsistencies that negatively affect the estimated level of underreporting. In addition, the END model presents estimated coefficient signs that are more consistent with the expected signs (see Table 3).

Table 5 **Estimated underreporting level for the FE, DEF, and END models**

	Mean	**Standard deviation**	**Minimum**	**Maximum**
FE	0.2120	0.0407	0.1999	0.4998
DEF	0.2500	0.0327	0.2034	0.3820
END	0.3273	0.1235	0.0077	0.8204

Source: Authors.

From 2004 to 2011, the END model's estimated average underreporting rate in Minas Gerais was 32.73%, with a standard deviation of 0.1235. If we consider the frequency distribution (Figure 4), 532 of the 827 Minas Gerais municipalities analysed had an underreporting level between 32% and 49%. Only 17 municipalities had an underreporting level above 65%, and 37 were below 10%. Figure 5 illustrates the estimated underreporting level's spatial distribution.

Figure 4 **Cumulative relative frequency distribution of the estimated underreporting in Minas Gerais's municipalities from 2004 to 2011**
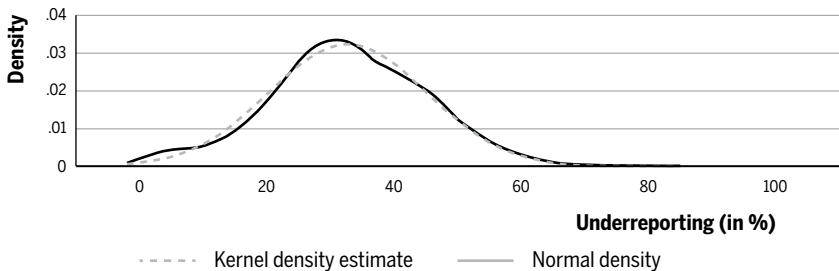
Source: Authors.

These results are close to those found in victimization surveys undertaken in Minas Gerais, although they adopted methodologies and sampling techniques that are distinct from those considered here. In the victimization survey conducted by the National Public Security Secretariat (2012), the average rate of underreporting in the state is 40.6%. The IBGE's (2009) victimization survey finds that the underreporting of robberies in Minas Gerais was 43%. The results from other victimization surveys applied in other Brazilian regions show significant underreporting rate differentials from the Minas Gerais mean. In the city of São Paulo, for example, the underreporting rate is measured at around 68% on average (Madalozzo and Furtado, 2011), and Cerqueira and Mello (2012) point out that the underreporting rates can reach as high as 80%.

Figure 5 **Spatial distribution of underreporting in Minas Gerais, mean values from 2004 to 2011**
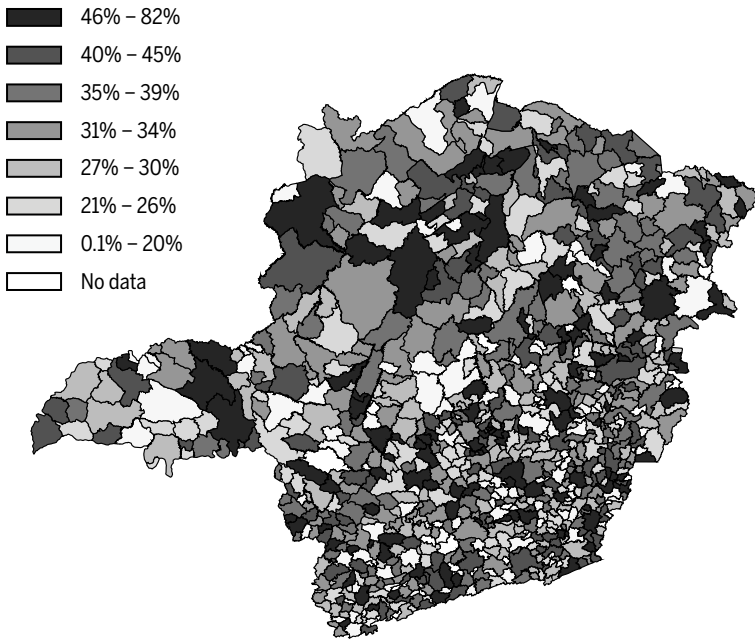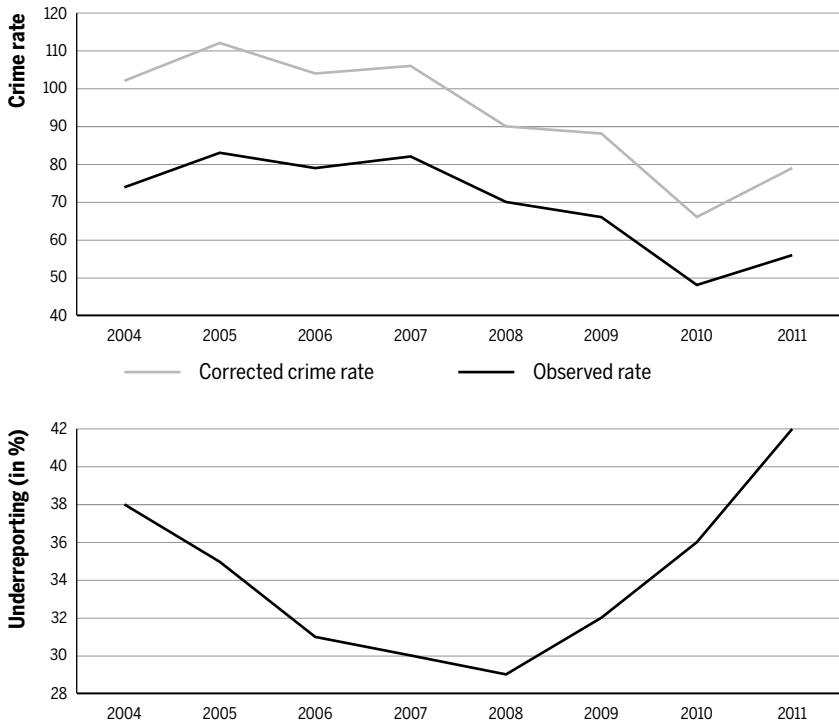

Source: Authors.

As previously discussed, crime underreporting creates bias when interpreting crime statistics. Figure 6 graphically depicts the official violent property crime rate, corrected for underreporting, and the underreporting rate as a percentage of the official crime rate from 2004 to 2011. The official violent property crime rate in Minas Gerais can be seen to have decreased by 32.5% from 2005 to 2011, from 83 to 56 crimes per 100,000 inhabitants; however, the corrected crime rate, achieved by adding unreported crimes to the official crime statistics, fell from 112 to 79 over the period, only a 29.5% decrease. All things being equal, the corrected crime rate should have fallen by 3% more and reached the official rate of 32.5%, but all things are not equal: the unreported crime rate rose as a percentage of the official rate from 2005 to 2011. The official crime rate at the start of 2011, while lower than that in 2005, overestimated the reduction in the violent crimes against property by 3%, a 9% deviation from the officially reported change. The bias caused by underreporting, a 9% error in this case, negatively affects

the allocation of resources needed to achieve the optimal outcome from the crime prevention and control policies in Minas Gerais.

Figure 6 **Official and corrected crime rates and the underreporting rate as a percentage of the official crime rate from 2004 to 2011**

Source: Authors.

Evidence of the bias in the official crime statistics generated by underreporting is verified in the study by Soares (2004a). When comparing the official statistics with victimization surveys for selected countries, the author finds that the official data report that, on average, only 2% of the population had suffered the crime of robbery, while the victimization surveys indicated that 25% of the population had experienced this kind of crime.

Analysing the graph on the right in Figure 6, it is apparent that underreporting is not constant over time, giving evidence of the need for caution in controlling for underreporting following the hypothesis that it is an unobservable and temporally invariant characteristic, as argued by

Justus and Kassouf (2008a). The variation in the rate of underreporting could be due to many factors, such as a change in the factors that affect the benefit–cost evaluation and in the actual or perceived efficiency of the relevant authorities.

Figure 6 shows that the official crime rate presented low variability from 2004 to 2007; however, the rate of underreporting fell from 38% to 30%. The combination of a relatively stable crime rate and a fall in the underreporting rate provides evidence of greater policing effectiveness that inhibited crime during the period. In 2008, the underreporting trend reversed, increasing to reach 42% in 2011. This suggests that victims stopped reporting crimes because the criminal justice system deteriorated and/or the fear of retaliation increased. Over the same 2008 to 2011 period, the official crime statistics show a decrease in crime. There appears to be a high likelihood that the official statistics from 2008 to 2011 were overly optimistic. If this was the case, it would reduce the net benefit from designing policies that depend on a proper interpretation of these crime statistics, but this can only be hypothesized. Because we do not have information about the type of goods stolen or whether violence was used in the act, we could also conclude that the police efficiency in combating crime in the 2004 to 2007 period increased the risk of punishment for criminal activity, thereby inhibiting criminal activity by changing the risk–reward characteristics: criminals sought targets who were less likely to report a robbery, either because the stolen item was of low value or because the targets were people who were unlikely to report a robbery to the authorities

Two tests were envisioned to verify the robustness of the study's underreporting estimates. In the first test, the estimates of underreporting were to be compared with victimization survey results. Unfortunately, this test was not carried out, because no municipal-level victimization surveys have been conducted in the study area.

In the second test, which was actually carried out, a placebo variable was inserted to replace the dependent variable in the END model's equation. The new dependent variable was the homicide rate per 100,000 inhabitants. In this exercise, a low underreporting rate was expected, since the occurrence of a homicide necessarily generates a record of the crime when the corpse is taken to the jurisdiction's *Instituto Médico Legal* for an autopsy. It should be noted that, due to the fact that the homicide rate is

null in many municipalities, one homicide was added to all the municipalities' data sets before logarithmic transformation to avoid the loss of observations. The results after this insertion indicated an average level of homicide underreporting in Minas Gerais from 2004 to 2011 of 4.47%, with a standard deviation of 0.0391. The minimum and maximum values obtained over the period were, respectively, 3.28% and 27.52%.

This homicide underreporting average of 4.47% can be contrasted with the results from a study by Cerqueira (2013). The objective of this author's study is to identify the existence of hidden homicides by examining the number of cases of death by undetermined causes that could be classified as homicides. As a main result, Cerqueira (2013) finds that an annual average of 8,600 homicides were not classified as homicides in Brazil from 1996 to 2010. In the case of the state of Minas Gerais, the author estimated that, from 2004 to 2010, the homicide rate was, on average, about 18% above the official rate.

5 Concluding remarks

The estimated average underreporting of 'violent crimes against property' in Minas Gerais, Brazil, from 2004 to 2011 is 32.7%. When this level of underreporting is taken into account, the crime rate in Minas Gerais fell 3% less than the official crime rate from 2005 to 2011, a 9% error in the official rate. If the official crime rate used to formulate policies during that period was not adjusted for this level of underreporting, the strategies designed to combat and prevent crime would be based on inaccurate statistics and quite possibly would be inappropriate strategies. Underreporting bias is a serious problem for the maintenance of public safety.

To verify the robustness of the study's results, a simple test was performed: the dependent variable, violent crimes against property, was replaced with the logarithm of the homicide rate. This change was made because the crime of homicide is one that supposedly presents a lower rate of underreporting. The results indicate average underreporting of homicides of 4.47% over the analysed period.

The methodology employed in this study led to the finding that the level of underreporting changes over time. This determination negates the hypothesis that is widely accepted in studies using regression analysis

that underreporting does not change over time and that the omission of this possibility from consideration does not affect the regression model's robustness.

Unfortunately, the study was limited by a lack of control variables. The stochastic frontier analysis method of estimation itself also has inherent limitations, such as the a priori definition assumed for the underreporting term's statistical distribution. Our study did not exhaust the subject, which merits study in greater depth. Crime underreporting affects the correct management of public safety.

References

Adamchik, V. A.; King, A. E. Labor market efficiency in Poland: a stochastic wage frontier analysis. *International Journal of Business and Finance Research*, v. 1, n. 2, p. 41-50, 2007.

Almeida, E. S. *Econometria espacial aplicada*. Campinas – SP. Alínea, 2nd ed., 2013. 498 p.

Almeida, E. S.; Haddad, E. A.; Hewings, G. J. D. The spatial pattern of crime in Minas Gerais: an explanatory analysis. *Economia Aplicada*, v. 9, n. 1, p. 39-55, 2005.

Araújo Jr., A. F.; Fajnzylber, P. *O que causa a criminalidade violenta no Brasil?* Uma análise a partir do modelo econômico do crime: 1981 a 1996. Discussion paper 162 – Universidade Federal de Minas Gerais, 2001. 88 p.

Arellano, M.; Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economics Studies*, v. 58, n. 2, p. 277-297.

Battese, G. E.; Coelli, T. J. Frontier production functions, technical efficiency and panel data: with applications to paddy farms in India. *Journal of Productivity Analysis*, v. 3, n. 1-2, p. 153-169, 1992.

Becker, G. S. Crime and punishment: an economic approach. *Journal of Political Economy*, v. 76, n. 2, p. 169-217, 1968.

Bezat, A. Comparison of the deterministic and stochastic approaches for estimating technical efficiency on the example of non-parametric DEA and parametric SFA methods. *Metody ilosciowe w badaniach ekonomicznych*, v. 10, n. 1, p. 20-29, 2009.

Brazilian Institute of Geography and Statistics (IBGE). Pesquisa Nacional por Amostra de Domicilio (PNAD), 2009.

Carvalho, J. R.; Lavor, S. C. Repeat property criminal victimization and income inequality in Brazil. *Revista EconomiA*, v. 9, n. 4, p. 87-110 2008.

Cerqueira, D. R. de C. *Mapa dos homicídios ocultos no Brasil*. Discussion paper 1848. Brasília–DF. Instituto de Pesquisa Econômica Aplicada (IPEA), 2013.

Cerqueira, D. R. de C.; Mello, J. M. P. de. *Menos armas, menos crimes*. Discussion paper 1721. Brasília–DF. Instituto de Pesquisa Econômica Aplicada (IPEA), 2012.

Chaudhuri, K. K.; Chowdhury, P.; Kumbhakar, S. B. Crime in India: specification and estimation of violent crime index. *Journal of Productivity Analysis*, v. 43, n. 1, p. 13-28, 2015.

Coelli, T. J. Recent developments in frontier modelling and efficiency measurement. *Australian Journal of Agricultural Economics*, v. 39, n. 3, p. 219-245, 1995.

Coelli, T.; Prasada Rao, D. S.; O'Donnel, C. J.; Battese, G. E. *An Introduction to Efficiency and Productivity Analysis*. 2nd ed., New York–NY. Springer, 2005. 356 p.

Cohen, L. E.; Kluegel, J. R.; Land, K. C. Social inequality and predatory criminal victimization: an exposition and test of a formal theory. *American Sociological Review*, v. 46, n. 5, p. 505-524, 1981.

Debreu, G. The coefficient of resource utilization. *Econometrica*, n. 19, p. 273-292, July 1951.

Goudriaan, H.; Wittebrood, K.; Nieuwbeerta, P. Neighborhood characteristics and reporting crime. *British Journal of Criminology*, v. 46, n. 1, p. 719-742, 2005.

Groot, W.; Ooterbeek, H. Stochastic reservation and offer wages. *Journal of Labour Economics*, v. 1, n. 3-4, p. 383-390, 1994.

Hausman, J. A. Specification tests in econometrics. *Econometrica*, v. 46, n. 6, p. 1251-1271, 1978.

João Pinheiro Foundation. Índice Mineiro de Responsabilidade Social. Belo Horizonte, 2015.

Justus, M. J. Dinâmica temporal da criminalidade: mais evidências sobre o 'efeito inércia' nas taxas de crimes letais nos estados brasileiros. *Revista EconomiA*, v. 10, n. 1, p. 169-194, 2009.

Justus, M. J.; Kassouf, A. L. Existe explicação econômica para o sub-registro de crimes contra a propriedade? *Economia Aplicada*, v. 12, n. 1, p. 5-27, 2008a.

Justus, M. J.; Kassouf, A. L. Estudos econômicos das causas da criminalidade no brasil: evidências e controvérsias. *Revista EconomiA*, v. 9, n. 2, 2008b.

Justus, M.; Kassouf, A. L. Evidence of the effect of wealth observed by criminals on the risk of becoming a victim of property crimes. *Revista EconomiA*, v. 14, n. 1, p. 1-14, 2013.

Justus, M. J.; Santos Filho, J. I. Convergência das taxas de crimes no território brasileiro. *Revista EconomiA*, v. 12, n. 1, p. 131-147,¬¬ 2011.

Justus, M.; Scorzafave, L. G. Underreporting of property crimes: an empirical economic analysis. *Economic Analysis of Law Review*, v. 5, n. 2, p. 271-284, 2014.

Kumbhakar, S. C.; Lovell, N. *Stochastic Frontier Analysis*. Cambridge, Cambridge University Press, 2000. 333 p.

Kumbhakar, S. C.; Wang, H.; Horncastle, A. P. *A Practitioner's Guide to Stochastic Frontier Analysis Using Stata*. Cambridge, Cambridge University Press, 2015. 290 p.

Kume, L. Uma estimativa dos determinantes da taxa de criminalidade brasileira: Uma aplicação em painel dinâmico. In: XXIII Encontro Nacional de Economia. ANPEC, João Pessoa, 2004. Available on: <http://econpapers.repec.org/paper/anpen2004/148.htm>.

Loureiro, A. O. F. Uma avaliação dos determinantes da criminalidade no Ceará. In: Encontro Economia do Ceará em Debate, Fortaleza, 2009.

Loureiro, A. O. F.; Carvalho Jr., J. R. A. O impacto dos gastos públicos sobre a criminalidade brasileira. In: XXXV Encontro Nacional de Economia, Recife, 2007.

Madalozzo, R.; Furtado, G. M. Um estudo sobre a vitimização para a cidade de São Paulo.

Revista de Economia Política, v. 31, n. 1, p. 160-180, March 2011.

National Public Security Secretariat. Pesquisa Nacional de Vitimização, Belo Horizonte, 2013. Available on: <http://www.crisp.ufmg.br/wp-content/uploads/2013/10/Relat%C3%B3rio-PNV-Senasp_final.pdf>.

Peixoto, B. T. *Determinantes da criminalidade no município de Belo Horizonte*. Universidade Federal de Minas Gerais, Belo Horizonte, 2003.

Peixoto, B.; Souza, L. G.; Lima, R. S. Uma análise sistêmica: vitimização e políticas de segurança em São Paulo. *Revista do Serviço Público*, v. 63, n. 2, p. 217-236, 2012.

Polachek, S. W.; Robst, J. Employee labor market information: comparing direct world of work measures of workers' knowledge to stochastic frontier estimates. *Journal of Labour Economics*, v. 5, n. 2, p. 231-242, 1998.

Prieto, C. G.; Román, A. M.; Domínguez, C. P. Actual and potential returns to schooling in Spain. *Education Economics*, v. 13, n. 4, p. 387-407, 2005.

Resende, J. P.; Andrade, M. V. Crime social, castigo social: o efeito da desigualdade de renda sobre as taxas de criminalidade nos grandes municípios brasileiros. *Estudos Econômicos*, v. 41, n. 1, p. 173-195 2011.

Sánchez, M. A. D. R. Gender and potential wage in Europe: a stochastic frontier approach. *International Journal of Manpower*, v. 32, n. 4, p. 410-425, 2011.

Sant'Anna, E.; Scorzafave, L. G.; Justus, M. Nonlinear relationship between income, age and criminal victimization in Brazil. *Revista EconomiA*, v. 17, n. 2, p. 185-198, 2016.

Shepard, R. W. *Cost and Production Functions*. Princeton University Press, Princeton, 1953.

Shikida, P. F. A.; Oliveira, H. V. N. Crimes violentos e desenvolvimento socioeconômico: um estudo sobre a mesorregião Oeste do Paraná. *Revista Brasileira de Gestão e Desenvolvimento Regional*, v. 8, n. 3, p. 99-114 2012.

Silver, E.; Miller L. L. Sources of informal social control in Chicago neighborhoods. *Criminology*, v. 42, n. 3, p. 551-83, 2004.

Soares, R. R. Crime reporting as a measure of institutional development. *Economic Development and Cultural Change*, Chicago, v. 52, n. 4, p. 851-71, 2004a.

Soares, R. R. Development, crime, and punishment: accounting for the international differences in crime rates. *Journal of Development Economics*, Chicago, v. 73, p. 155-184, 2004b.

Teixeira, E. C. *Dois ensaios acerca da relação entre criminalidade e educação*. Escola Superior de Agricultura Luiz de Queiroz – Esalq/USP, Piracicaba–SP, 2011.

Terza, J. V.; Basu, A.; Rathouz, P. J. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of Health Economics*, v. 27, n. 3, p. 531-543, 2008.

Tolentino, N. F.; Diniz, A. M. A. A distribuição geográfica do tráfico de drogas em Belo Horizonte e suas correlações espaciais. *Revista do Instituto de Ciências Humanas*, v. 9, n. 12, p. 66-73 2014.

Tran, K. C.; Tsionas, E. G. GMM estimation of stochastic frontier model with endogenous regressors. *Economics Letters*, v. 118, n. 1, p. 233-236, 2013.

Tsyzler, M. *Econometria espacial*: Discutindo medidas para a matriz de ponderação espacial.

Fundação Getúlio Vargas, São Paulo–SP, 2006.

Warner, B. Directly intervene or call the authorities? A study of forms of neighborhood social control within a social disorganization framework. *Criminology*, Oxford, v. 45, n. 1, p. 99-129, 2007.

About the authors

Gustavo Carvalho Moreira - gustavocmoreira@usp.br
Universidade Federal de São João del Rei – UFSJ, São João del Rei, Minas Gerais, Brasil.
ORCID: https://orcid.org/0000-0001-9526-0267.

Ana Lucia Kassouf - anakassouf@usp.br
Professora Titular do Departamento de Economia da Escola Superior de Agricultura Luiz de Queiroz – Esalq/Universidade de São Paulo, Piracicaba, São Paulo, Brasil.
ORCID: https://orcid.org/0000-0002-1931-5972.

Marcelo Justus - mjustus@unicamp.br
Professor e pesquisador do Instituto de Economia da Universidade de Campinas – IE/Unicamp, Campinas, São Paulo, Brasil.
ORCID: https://orcid.org/0000-0001-8660-1779.

About the article

APPENDIX

Table A1 **Descriptive statistics of the variables selected for the empirical model**

Variable		Mean	E.S.	Min.	Max.
police	*Total*		88.81	0.00	1,222.13
	Between	142.39	83.18	25.36	771.50
	Within		31.22	-383.22	649.69
income	*Total*		588.06	79.23	7,057.84
	Between	1,992.23	496.78	1,003.39	6,769.22
	Within		315.09	-32.23	5,755.11
GDPpercapita	*Total*		10,326.38	982.9	283,298.20
	Between	7,821.54	8,236.60	2,185.42	129,076.10
	Within		6,234.22	-87,397.49	234,431.90
young	*Total*		2,389.90	16,493.26	45,506.51
	Between	25,498.18	2,083.49	19,223.92	36,710.41
	Within		1,172.65	17,737.01	38,774.15
urban	*Total*		19.38	14.60	100.00
	Between	66.52	18.34	16.86	100.00
	Within		6.26	2.33	134.02
vulnerability	*Total*		100.09	4.32	550.31
	Between	162.53	74.73	27.69	363.97
	Within		66.62	-94.61	412.80
Wcrime	*Total*		13.82	39.02	219.97
	Between	69.89	6.87	56.52	149.80
	Within		11.99	10.78	140.07
crime(-1)	*Total*		109.80	0.00	1,694.71
	Between	69.86	98.77	0.00	1,149.69
	Within		48.06	-442.58	714.54

Note: Within: variance between municipalities; between: average variation in time between municipalities. Descriptive statistics were generated before logarithmization.