**Features of crime in Brazil: an approach based on decision trees algorithms** [r]

Tauã Magalhães Vital[°]

Daniel Morais de Souza[°]

Admir Antonio Betarelli Junior**

Rogério Silva de Mattos**

**Abstract**

Violence is one of the biggest challenges facing Brazilian society. The number of homicides in the country has been growing heterogeneously throughout its territory since the 1980s. Recent studies have analyzed the impacts of public policies on crime. However, economic and social conditions may be far more important to explain crime. This paper aims to investigate the features that contribute the most to characterize crime in Brazil. Employing an approach based on decision trees algorithms we were able to analyze a significant amount of variables that can impact crime. The main findings show that housing conditions, demographic density, lack of religion and the number of homicides in neighboring regions are significant to explain the 2016 homicide rate in Brazilian cities.

**Key-words:** crime; machine learning; decision trees.

June 2020

[r] This paper was supported by CNPq and FAPEMIG under the grant "research project", and CAPES under the grant "PhD Scholarship".

° PhD in Economics at the Federal University of Juiz de Fora (UFJF); **Professors of the Economics Department, UFJF.

## 1. Introduction

Since the 1980s, crime has been rising in Brazil, making the country one of the most violent in the world. According to the Ministry of Health's Mortality Information System, DATASUS[1], in 2017, Brazil had a rate of 30.7 willful killings per 100,000 inhabitants. Given this scenario, crime is one of the biggest problems facing the country.

Violence is a social, economic and political problem, and is directly related to economic conditions, thus affecting the potential development of nations (Fajnzylber and Araujo Jr 2001). Consequently, the active participation of government spheres in the orientation of public security policies is necessary. Estimates indicate that the decrease in human capital stock resulting from homicides are nearly 2.25 billion dollars to the detriment of Brazil for 2001 (Carvalho and Lavor 2008).

There is a recent literature on causal inference of public policy in econometrics, many related to crime (Wilson, Petersilia, and Wilson 2002; T. M. Vital 2018; Cabral 2016; Castro, n.d.). The evaluation of public policies should be guided in order to be able to determine its causal effect, isolating its impact from other characteristics present in econometric models.

After its redemocratization, the Brazilian state has undergone drastic economic and social changes that may have impacted the decision of an individual to commit a crime or not. The availability of data has made possible for researchers to devote some effort exploring the main determinants of crime in Brazil (Daniel et al. 2017; Cerqueira 2014; Almeida, Haddad, and Hewings 2005; Chioda, De Mello, and Soares 2016).

An intelligent system that can predict the place and time at which an offense will occur could become a good apparatus in combating crime. However, due to the randomness in space and time, linked to the incidence of crimes, the creation of such a system proves to be a tough task, encouraging several researchers to study this phenomenon.

In 2010, UCLA researchers developed a pioneering system, Predpol, that has been used by Los Angeles police and in several US cities since them. From this, the concept of predictive police was developed, whereby an intelligent system is updated from data constantly predicting crime hot spots and allowing effective allocation to these areas. However, such systems have been criticized as to the bias caused by it. Evidence points to a reduction in crime in hot spots and surrounding areas linked to predictive police practices, but there are no studies related to the impact of the use of such police practice on minority communities (Braga, Papachristos, and Hureau 2014).

The evolution of computational technology has allowed the diffusion of machine learning techniques in economics related applications. Since the 1990's methods like bagging, boosting and random forest, that are based in the repetitions of the experiments, and have a big computational cost are now feasible to implement in a short period of time. Most studies using learning-based techniques are used for prediction, however, they can be an important ally in robustness of results (Athey 2015), especially decision tree-based machine learning methods..

This paper uses decision tree-based algorithms to assess the main features of crime in Brazil. Bagging, random forest and boosting methods were used to improve robustness to the results. Given that such methods decrease the variance of the simple decision tree algorithm. One of the advantages of decision trees algorithms is the possibility of analyzing a bigger number of variables than conventional econometrics. Since to the best

[1] The homicide rate was obtained using the sum of ICD-BR-10 categories X85-Y09 (assaults) and Y35-Y36 (legal interventions and war operations), available at http: //tabnet.datasus. gov.br/cgi/tabcgi.exe?sim/cnv/ext10br.def.

of our knowledge there is no similar study employing machine learning algorithms to explore the features of crime in Brazil, this paper fills this gap in the crime literature. We used data of the 5570 Brazilian municipalities for 2016, having as dependent variable the homicide rate per 100 thousand inhabitants and as features several variables based on the economic model of crime of Becker (1968).

Besides this introductory section, this paper has five more sections. The second presents some of the most relevant work related to the economics of crime. The third section discusses the methodology used as well as a description of the data. The fourth presents the results found. Finally, the last section discusses them, as well as possible policy implications of the results.

## 2. Economics and crime

Ecological theories seeks to explain variations in crime rates through the differing incentives faced by the individual (Kelly 2000). Those incentives are influenced by different environments, better economic and social local conditions can result as deterrent effect on aggregated crime data. The economic approach came from the studies of Fleischer (1963, 1966), Ehrlich (1967) and Becker (1968).

In the economic crime model of Becker (1968) is assumed the premise of rationality of criminals. The individuals seek to maximize their expected utility. In other words, the individual makes the decision whether or not to commit the crime based on an analysis of the expected costs and benefits of the criminal practice. The individual weighs his or her decision by comparing the expected returns from committing a crime with the expected returns from legal alternatives. Therefore, crime is analyzed as any other economic activity. As demonstrated in a recent study by Furqan and Mahmood (2020), education, GDP per capita, demographic density and unemployment have an influence in the utility function about the decision of an individual to commit a crime.

Most of the empirical tests of economic crime theory are focused in the deterrence or incapacitation effect of police (Levitt 1997; 2004; Murray, de Castro Cerqueira, and Kahn 2013; Wilson, Petersilia, and Wilson 2002; Levitt 2017). In Brazil, Manso and Dias (2018), and Dias (2013) show evidence that organized crime in Brazil has a very strong contagion effect on young people from the outskirts, influencing them to enter the world of crime. The Brazilian prison system is extremely ineffective in isolating the convicted criminal from society, allowing him to maintain contact with outside criminal organizations. The presence of organized crime in peripheral regions, such as the First Capital Command (PCC) in São Paulo and the Red Command (CV) in Rio de Janeiro, makes young residents of these more state-relegated regions see in the world of crime. the way out for a life improvement.

A theory that aims to explain the behavior of the criminal must take into account the motivations of individual behavior, as well as the effect of the environment in which the individual finds himself, paying attention to the spatial and temporal distribution of crime (Kelly 2000).

As proposed by Vital, De Souza, and Faciroli (2020) the decision to commit or not a crime by an individual $i$ (such that, $i = 1, 2, \dots, n$ , where n is the population size) is influenced by an intrinsic factor and the factors external to the individual. Thus, the decision to commit a felony at time $t$ and location $s$, $Y_{i,s,t}$, may be expressed by the following expression:

$$Y_{i,s,t}(t) = \lambda_i(t) + f(X_{s,t}) \tag{1.1}$$

The decision of the individual to commit the crime depends on its intrinsic characteristics, $\lambda_i$, represented by static factors such as neurological profile, personality, age, gender, psychopathy, among others. Also, dynamic factors, $f\left(X_{s,t}\right)$, such as drug use, unemployment, peer effect, educational level and other socioeconomic factors. Note that the dynamic factors are those in which public policies can have influence. In this model the functional form of $f(X_{s,t})$ is not assumed to be linear as most of the econometric papers does. In this paper we aim to address this issue allowing the tree-based algorithm to find the best fit for the model.

Note that in contrast to the model proposed by Becker (1968) the decision of the individual to commit the crime is not related to the utility derived from the income provided by the crime. As an example, individuals with a high degree of psychopathy, $\lambda_i(t) > f(X_{s,t})$, would commit the crime regardless of whether or not they receive monetary compensation.

## 3. Empirical estrategy and dataset

### 3.1 Data

We made the choice of variables incorporated in the analysis based on the model proposed by Becker (1968) and already known in the economic crime literature. The dependent variable is the homicide rate for the 5,564 Brazilian municipalities for 2016. Due to small under reporting of this type of crime, homicide is the best proxy for violence available for the entire country.

The independent variables are: population density, GDP per capita, percentage of the population without religion, ratio of the 20% richest of the population to the 40%. poorest, percentage of young men aged 15-24, percentage of head of households without basic education with children under 15, police expenditure in Brazilian reais by 100,000 inhabitants, number of local police officers by one hundred thousand inhabitants, number of local police officers that carry firearms by one hundred thousand inhabitants, and the IBEU-Municipal indicators (RIBEIRO and RIBEIRO 2016).

The displacement effect of crime in Brazil was already identified by Almeida, Haddad, and Hewings (2005). Taking that into account we used the procedure proposed by Baumont (2009) and created a spatial lagged variable of homicide rate.

The IBEU-Municipal indexes are based on the 2010 IBGE census data for Brazil using principal component analysis. These indicators represent the well-being of the population resulting from urban conditions. In many ways the welfare is intrinsically related to the crime rate. A city where the population feels ignored by the state tends to resort to crime, which is one of the reasons for the high crime rate in suburban areas. To the best of our knowledge this paper is the first to incorporate this data in the crime analysis.

The urban mobility index is based on the commuting time that an individual spends between home and work. It can be considered as a proxy for transport infrastructure. The crime displacement effect leads us to the hypothesis that municipalities with low urban mobility (good transport infrastructure) would offer greater police effectiveness and therefore a lower crime rate. Also, commuting time is directly related to the utility function of the individual in committing a crime, since the individual's remuneration is affected by it.

Urban environmental conditions index takes into account three factors: afforestation, open sewage and accumulated garbage around households. The variable for collective services reflects the population's access to basic services such as treated water, garbage collection and energy. The urban infrastructure index addresses street lighting, paving, and wheelchair access on the premises. Lastly, housing conditions index reflects the situation of household density and housing precariousness. Our testing hypothesis is a municipality with precarious urban environmental conditions would present a high crime rate. Ferreira, Bastos, and Betarelli Junior (2019) recently did a study using factor analysis and QCA reached the conclusion that a lower social coactivity is associated with higher homicide rates, emphasizing the importance of the inclusion of these control variables in our study.

Recent literature (Cabral 2016; Pereira Filho and De Sousa 2018; T. M. Vital 2018) signal a possible deterrent effect linked to the implementation of municipal guards in recent years. However, it is conjectured that this impact is also associated with differences between municipal guards in Brazil, some of which may prove more effective in reducing crime than others. Thus, the analysis included the time of existence of such apparatus as a proxy for effectiveness, and the use of lethal weapons. Information about the municipal guards was taken from the 2014 MUNIC[2] from IBGE.

[2] Munic is a survey conducted by IBGE on the profile of Brazilian municipalities. Available at http://downloads.ibge.gov.br/downloads_estatisticas.htm

Table 1. Definition of the variables and descriptive statistics

| Variable | Description | Source | Mean | Std. Deviation | Max. | Min | Obs. |
|---|---|---|---|---|---|---|---|
| homicide | Homicides by one hundred thousand people | DATASUS/Ministry of Health | 22.140 | 23.750 | 234.680 | 0.000 | 5507 |
| whomicide | Spatial lagged homicide rate | DATASUS/Ministry of Health | 22.21 | 14.152 | 126.44 | 0.000 | 5507 |
| expermunpol | Time of existence of Municipal Guards | Munic 2014 | 2.78 | 7.865 | 86.000 | 0.000 | 5507 |
| bolsafam | Percentage of population benefited by the program “Bolsa Família” | PNAD/IBGE | 107.710 | 23.619 | 350.000 | 12.280 | 5507 |
| gdp | GDP per capita | Census/IBGE | 9749.000 | 14113.25 | 296885.0 | 2270.0 | 5507 |
| demodensid | Population per km² | Census/IBGE | 108.421 | 572.244 | 13024.562 | 0.131 | 5507 |
| youngpop | Percentage of 15 to 29 to years old, male of the total population | Census/IBGE | 13.284 | 1.506 | 37.054 | 8.099 | 5507 |
| noreligion | Percentage of the total population who does not have a religion | Census/IBGE | 5.415 | 4.977 | 54.234 | 0.055 | 5308 |
| femalehead | Percentage of single mothers without basic education in the total population | Census/IBGE | 20.010 | 10.300 | 77.590 | 0.000 | 5507 |
| inqeuality | Ratio between the 20% richer of the population and the 40% poorest | Census/IBGE | 9.591 | 6.400 | 179.49 | 0.000 | 5507 |
| urbanmob | Urban mobility index | IBEU/IBGE | 0.938 | 0.062 | 1.000 | 0.009 | 5507 |
| habcond | Habitation conditionals index | IBEU/IBGE | 0.857 | 0.071 | 0.988 | 0.459 | 5507 |
| urbancond | Urban condition index | IBEU/IBGE | 0.837 | 0.131 | 1.000 | 0.242 | 5507 |
| urbanser | Urban public services index | IBEU/IBGE | 0.718 | 0.174 | 1.000 | 0.263 | 5507 |
| urbaninfra | Urban infrastructure index | IBEU/IBGE | 0.513 | 0140 | 0.926 | 0.081 | 5507 |
| infocrim | Dummy variable of Infocrim | Secretary of Public Security/SP | 0.009 | 0.090 | 1.000 | 0.000 | 5507 |
| upp | Dummy variable of UPP | Secretary of Public Security/RJ | 0.0002 | 0.013 | 1.000 | 0.000 | 5507 |
| ronda | Dummy variable of Ronda Quarteirão | Secretary of Public Security/CE | 0.001 | 0.0301 | 1.000 | 0.000 | 5507 |
| pacto | Dummy variable of Pacto pela Vida | Secretary of Public Security/PE | 0.034 | 0.180 | 1.000 | 0.000 | 5507 |
| polexpenditure | Police expenditure in brazilian reais by one hundred thousand people | FINBRA/Ministry of Finance | 3.696 | 14.308 | 313.118 | 0.000 | 5102 |
| munpol | Number of guards by one hundred thousand people | Munic 2014 | 2012 | 25.040 | 80.284 | 1762.950 | 0.000 |
| armedmunpol | Number of armed guards by one hundred thousand people | Munic 2014 | 2012 | 3.245 | 25.300 | 701.884 | 0.000 |
| quality | General index of quality of public services | IBEU/IBGE | 0.773 | 0.083 | 0.951 | 0.444 | 5507 |

Source: elaborated by the authors.

### 3.2 Empirical approach

#### 3.2.1 Machine Learning Vs Traditional Econometrics

In this section we begin our discussion introducing what are the differences in the concerns and goals between traditional econometrics and the machine learning literature. Then we focus on the tree-based methods, which are the class of methods that we used.

The tradicional approach in econometrics is to specify a target, often a parameter of a statistical model, that is a functional of a jointed distribuition of the data (Athey and Imbens, 2019). The main goal is to estimate the parameter of interest by choosing the parameter values that best fit the data using an objective function like sum of squared errors or likelihood. The focus of traditional econometrics is on the quality of the estimators of this target, because usually it is important to understand an object like $E(Y|X)$ in order to perform exercices of evaluating the impact of changing one covariate while holding others constant (Wooldridge, 2010).

In contrast, the machine learning literature focuses on developing algorithms to be applied to any datasets, with main areas being prediction (regression), classification, and clustering or grouping tasks (Athey, 2018). Burkov and Lutz (2019) defined Machine Learning as the process of solving a practical problem by gathering a dataset and algorithmically building a statistical model based on that dataset.

We can divide the machine learning techniques into two main branches, supervised and unsupervised learning. The goal of unsupervised learning is to create clusters of observations that are similar in terms of their covariates, and thus can be interpreted as "dimensionality reduction" (Athey, 2018). Supervised learning algorithms seek functions that predict well out of sample, by taking a loss function $\hat{L}(\hat{y}, y)$ as an input and search for function $\hat{f}$ that has low expected prediction loss $E_{(y,x)} = [L(\hat{f}(x), y)]$ on a new observation from the same distribution (Mullainathan and Spiess, 2017).

There are a variety of machine learning methods for supervised learning, such as regularized regression (rigde, lasso, elastic net, least angle and partial least squares), basis expansions and regularizations, kernel smoothing methods, regression trees, neural nets, support vector machines, matrix factorization, model averaging and many others. Perhaps the easiest method to interpret among these is the regression trees, because the model result are logical structures that can be understood with no statistical knowledge (Lantz, 2015).

Regression trees involves stratifying the predictor space into a number of sample regions in order to make a prediction for a given observation (James et al., 2013).Some advantages of this method are: natural handling of data of mixed type, handling of missing values, robustness to outliers in covariates space, computational scalability and ability to deal with irrelevant covariates (Friedman, Hastie, and Tibshirani 2001).

The major problem presented by regression trees is that they suffer from high variance. One way of deal with this problem is to produce multiple trees and then combine all to yield a single consensus prediction (James et al., 2013). There are three widespread methods that do this: bagging, random forests and boosting. In a recent study, Ettensperger (2019) evaluated a different set of machine learning algorithms in social-economic data. The results presented by the author showed that the random forest algorithm is more suited to explain the relationship between the variables, given that the latter is not linear. In the next subsections we will present the *tree-based* methods previously mentioned.

### *3.2.2 Decision Trees*

The decision tree method is closely related to the concept of entropy. An informal and simple definition of entropy is as a measure of disorder or uncertainty. Since the goal of most machine learning algorithms is to reduce uncertainty, entropy reduction and therefore information gain is one of the pillars of decision tree algorithms.

According to Hastie (2013), the decision trees algorithm needs to decide on the splitting variables and split points, and also what shape the tree should have. For easy interpretation, it is used to divide the predictor space into high-dimensional rectangles, or *boxes*. The decision tree method partition the covariates space in *M* distinct regions (or *boxes*) $(R_1, R_2, \dots, R_M)$, and than fit a simple model (like a constant) in each one $c_m$:

$$f(\boldsymbol{X}) = \sum_{m=1}^{M} c_m \boldsymbol{I}(\boldsymbol{X} \in R_m) \tag{3.1}$$

where $\boldsymbol{I}(\boldsymbol{X} \in R_m)$ is the information matrix in the region $R_m$. Notice that the best fit considering sum of squares as the minimization criterion is just the average of $y_i$ in region $R_m$, that is $\hat{c}_m = ave(y_i|x_i \in R_m)$. It is computationally infeasible to find the best binary partition in terms of minimum sum of squares, so, another method must be used. The *top-down* greedy algorithm is commonly used to solve this kind of problem (JAMES *et al*, 2013).[3] In order to perform recursive binary partition, consider a splitting variable *j* and a split point *s*, and define the pair of half-planes:

$$R_1(j,s) = \{\boldsymbol{X}|X_j \leq s\} \; e \; R_2(j,s) = \{\boldsymbol{X}|X_j > s\} \tag{3.2}$$

Then we seek the splitting variable *j* and split point *s* that solve:

$$min_{j,s}\left[\sum_{i:x_i \epsilon R_1(j,s)} (Y_i - c_1)^2 + \sum_{i:x_i \epsilon R_2(j,s)} (Y_i - c_2)^2\right] \tag{3.3}$$

For any choice *j* and *s*, the inner minimization is solved by:

$$\hat{c}_1 = ave(Y_i|X_i \in R_1(j,s)) \text{ e } \hat{c}_2 = ave(Y_i|X_i \in R_2(j,s)) \tag{3.4}$$

Having found the best split, we partition the data into the two resulting regions and then repeat the splitting process. However, in this step, instead of splitting the entire predictor space, we split one of the two previously identified regions. The process continues until a stopping criterion is reached. The determination of the split point *s* for each splitting variable can be done very quickly by scanning through all the covariates, enabling the determination of the best pair (*j,* s) (Hastie, 2013).

As pointed out by James *et al.* (2013), decisions tree may produce good predictions on the training set, but is likely to overfit the data, leading to poor test set performance. This is because the resulting tree might be too complex. A smaller tree with

[3] A greedy algorithm is a type of algorithm that seeks to find the best solution segment by segment. In the case of the decision tree, it looks for the segment that offers the highest information gain and therefore, the lowest entropy.

fewer splits might lead to lower variance and better interpretation at the cost of a little bias. Tree size is a tuning parameter governing the model's complexity, and the optimal tree size must be chosen from the data (Hastie; 2013).[4]

The best strategy to create a small tree is to grow a very large tree $T_0$, stopping the splitting process only when some minimum node size is reached, and then prune it back by deleting bottom nodes through a process of statistical estimation (Torgo, 2003). One method that can be used to prune a regression tree is the Cost-complexity pruning. This process consider a sequence of trees indexed by a nonnegative tuning parameter α, such that for each value of α exist a corresponding subtree $T \subset T_0$ represented by (3.5).

$$C_\alpha(T) = \sum_{m=1}^{|T|} \sum_{i:x_i \in R_m} (y_i - \hat{y}_{R_m})^2 + \alpha|T| \tag{3.5}$$

where $|T|$ represent the number of terminal nodes of the tree $T$, $R_m$ is the rectangle corresponding to the *m*th terminal node, and $\hat{y}_{R_m}$ is the predicted response associated with $R_m$. The idea of Cost-complexity prune is to find the subtree $T_\alpha \subset T_0$ that minimizes $C_\alpha(T)$ for each α. The tuning parameter α controls a trade-off between tree size (subtree's complexity) and its fit to the training data. When $\alpha = 0$, the subtree *T* will simply equal to the full tree $T_0$. As $\alpha$ gets larger, the subtree gets smaller. We can choose α using cross-validation by minimizing the cross-validated sum of squares (Hastie, Tibshirani, and Friedman, 2009)

As mentioned before, one major problem presented by decision trees is that they suffer from high variance, i.e., different training sets can yield to discrepant results. The reason for this inconsistency is the propagation of the errors in the top splits to all of the splits below it. One way to solve this problem is to use approaches that involves producing multiple trees which are then combined to yield a single consensus prediction (JAMES *et al*, 2013). There are three approaches consolidated in the machine learning literature that uses this logic, i.e., averages many trees to reduce this variance: *bagging*, *random forest* and *boosting*. The first and the last one are the two principal ensemble learning methods[5]. These methods will be explained in the next section. However, we restrict our discussion of these methods to the context of decision trees.

#### *3.2.3.1 Bagging*

To understand how the bagging method works, first we must understand the concept of bootstrap. Bootstrap is a technique that generate multiple samples from a single sample by drawing instances from the original sample with replacement (Alpaydin, 2009). Bagging, or bootstrap aggregation, is a general-purpose procedure for reducing the variance of a statistical learning method. Consider a decision tree prediction $\hat{f}(x)$ with covariates vector *x*. If we take *B* bootstrap samples, for each bootstrap $Z^b$, $b = 1, 2, \dots, B$, we could calculate $\hat{f}^1(x)$, $\hat{f}^2(x)$, ..., $\hat{f}^B(x)$. Bagging method averages this prediction over a colection of bootstrap samples. The bagging estimate is defined by:

$$\hat{f}_{bag}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}^b(x) \tag{3.6}$$

[4] The tuning parameter serves to control the trade-off between accuracy and variance.

[5] Ensemble learning focuses on training a large number of low-accuracy models and then combining the predictions given by those weak models to obtain a high-accuracy model (BURKOV; 2019).

The trees from bootstrap samples are grown deep, and are not pruned, which produces trees with high variance, but low bias. Averaging these $B$ trees reduces the variance. Bagging has been showed impressive improvements in accuracy by combining together hundreds or even thousands of trees into a single procedure (JAMES *et al.*, 2013).

Unfortunately, this improvement in accuracy comes with expensive interpretability cost. When we combining a large number of trees the resulting model is no longer a single tree, and it can be hard to interpret. Nevertheless, it's possible to identify the importance of each predictor using the residual sum of squares from the collection of bagged trees. We must gather the total amount that the residual sum of squares is decreased due to splits over a given predictor, averaged over all $B$ trees. A large value indicates an important predictor. In machine learning literature this measure is commonly called as variable importance.

#### *3.2.3.2 Random Forest*

Random forest is an enhancement of the bagging method that builds a collection of *de-correlated* trees and then averages them. The only difference between random forest and bagging procedures is that each time a split in a tree is considered, a random sample of $m$ predictors is chosen as split candidates from the set of $p$ predictors (JAMES *et al*, 2013)[6]. The random forests algorithm is used when the correlation between variables is high so that the bagging algorithm is not efficient in reducing the variance between samples.

According to Hastie (2013), to grow a random-forest tree $T_b$ to the bootstrapped data $b$ you must recursively repeat the following steps for each terminal node of the tree until some criterion is achieved: *i)* select $m$ variables at random from the $p$ variables; *ii)* pick the best variable and split point among the $m$ predictors and; *iii)* split the node into two nodes. After $B$ trees $\{T(x;\Theta_b)\}_1^B$ are grown; being $\Theta_b$ the $b$th random forest tree in terms of split variable and cutpoints at each node; the random forest regression predictor is:

$$\hat{f}_{rf}(x) = \frac{1}{B}\sum_{b=1}^{B} T(x;\Theta_b) \tag{3.7}$$

The idea of random forest algorithm is to reduce the variance of bagged trees by eliminating the correlation between them. As pointed by Hastie (2013), the bias of bagged trees is the same as that of individual bootstraped trees, and then, the only way of improving model performance is through variance reduction. One of the disadvantages of using the random forest algorithm is the computational power required, often causing a high processing time.

#### *3.2.3.3 Boosting*

As Bagging and Random Forest methods, boosting also involves combining multiple trees in order to establish a single predictive model. Boosting works in a similar way of bagging, except that the trees are grown sequentially, i.e., each tree is grown using the information from previously grown trees (Burkov and Lutz, 2019). Each new tree would be different from the previous ones in the sense that it tries to fix the errors which previous trees make.

---

[6] Typically values for $m$ are $\sqrt{p}$ or even as low as 1.

According to James *et al.* 2013, to perform boosting algorithm we first have to set up $\hat{f}(x) = 0$ and $r_i = y_i$. Then, we repeat the following steps for $b = 1, \ldots, B$: *i)* fit a tree $\hat{f}^b$ with $d$ splits to the training data $(X, r)$; *ii)* update $\hat{f}$ by adding in a shrunken version of the new tree, i.e. $\hat{f}(x) = \hat{f}(x) + \lambda \hat{f}^b(x)$ and; *iii)* update the residuals, $r_i = r_i - \lambda \hat{f}^b(x)$. Performed the *B* steps, the boosting regression predictor is:

$$\hat{f}_{boost}(x) = \frac{1}{B} \sum_{b=1}^{B} \lambda \hat{f}^b(x) \tag{3.8}$$

After the first cycle, the boosting algorithm fits a decision tree to the residuals from the current model rather than the outcome. These trees can have just a few splits determined by the parameter *d* in the algorithm. By fitting trees with small values of *d* and λ to the residuals we slowly improve $\hat{f}$ in areas where it does not perform well (James et al., 2013). The number of splits *d*, the shrinkage parameter λ and the number of trees *B* are the tuning parameters of boosting algorithm.

## 4. Results

The data was splitted between 75% of the observations for training and the remaining for testing. The dependent variable is the homicide rate due to small under reporting associated with this kind of crime, allowing us to get unbiased results due to measurement error.

Firstly, we adjusted a simple regression tree, and then improved its performance using bagging, boosting and random forest. The fit of the model to the data was based on the mean squared error (MSE). The decision tree algorithm presented an MSE equal to 457.2 in the training data set, whereas in the test data set 467.9. Tree pruning was tested and no performance improvement was found, that is the most complex tree was selected using cross-validation.

Using the bagging method to improve performance resulted in an MSE equal to 408.7 in the test set, while the boosting showed an improve 406.5. The random forest method presented an MSE of 398.9 in the data test set, that is, using this method improves the adjustment of the results found by the decision tree method. The latter showed the best fit in the testing set.

From the 24 variables used in the analysis, only five were selected in the random forest as important to construct the final tree: housing conditions (*habcond*), percentage of the population with no religion (*noreligion*), populational size (*population*), demographic density (*demodensid*), and the spatial lagged homicide rate (*whomicide*). The choice of the variables is based on two criterias. The first is based on the decay of mean accuracy of the predictors in out of the bag samples when a given variable is withdrawn from it. The second is a measure of the decay of impurity that results in a tree partition when a particular variable is removed, based on the sum of the squares of the residues. Figure 1 presents the results of random forest importance.

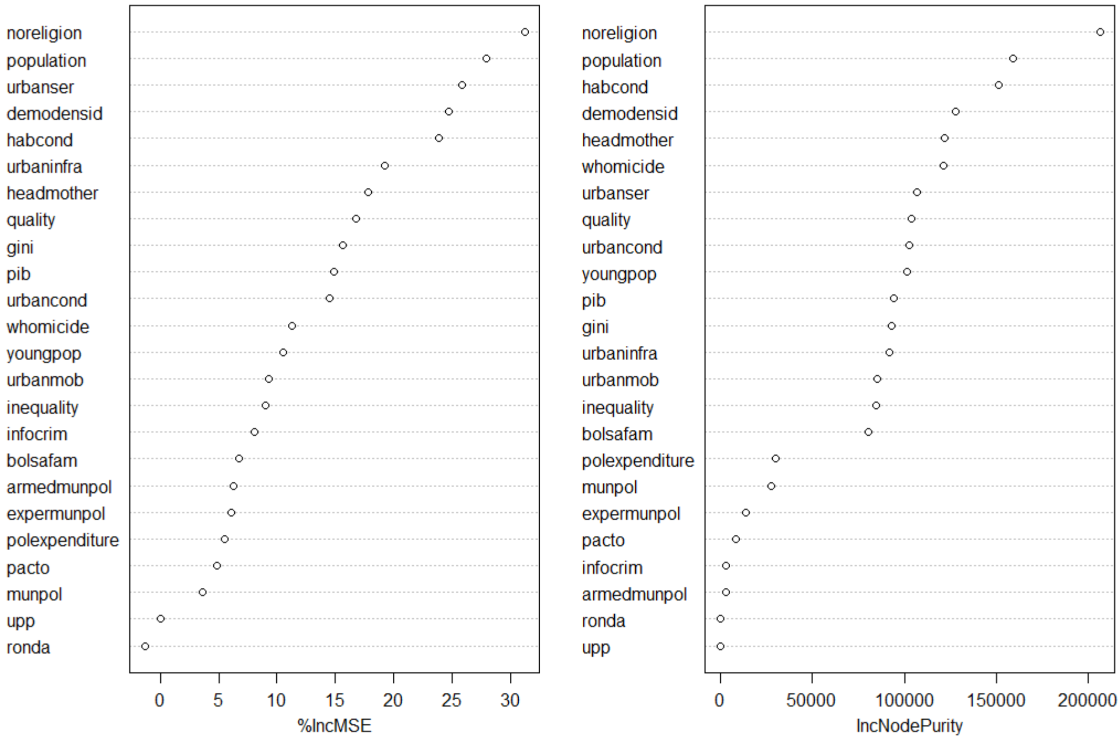


Figure 1: Importance of the variables in random forest.
Source: elaborated by the authors

From the random forest analysis it is possible to infer that all the variables associated with police, such as the public expenditure in police, the number of local police officers by 100,000 inhabitants, the experience of the local police measured in years of existence, the dummy indicating carrying firearms by them, and other dummy variables related to public policies to fight crime (presence of UPPs in Rio de Janeiro, the program Pacto pela Vida in Ceará State, Ronda Quarteirão in Pernambuco State, and Infocrim in São Paulo), did not improve the performance of the algorithm. And variables associated with social and economic conditions are in the top of random forest's importance of variables. This results points in a direction that the high crime rate in Brazil are more linked to poor social conditions than the efficiency of police. Using the results of random forest, we picked the five variables that improved the accuracy and we constructed the final tree for all the cities in Brazil. Figure 2 presents the results.

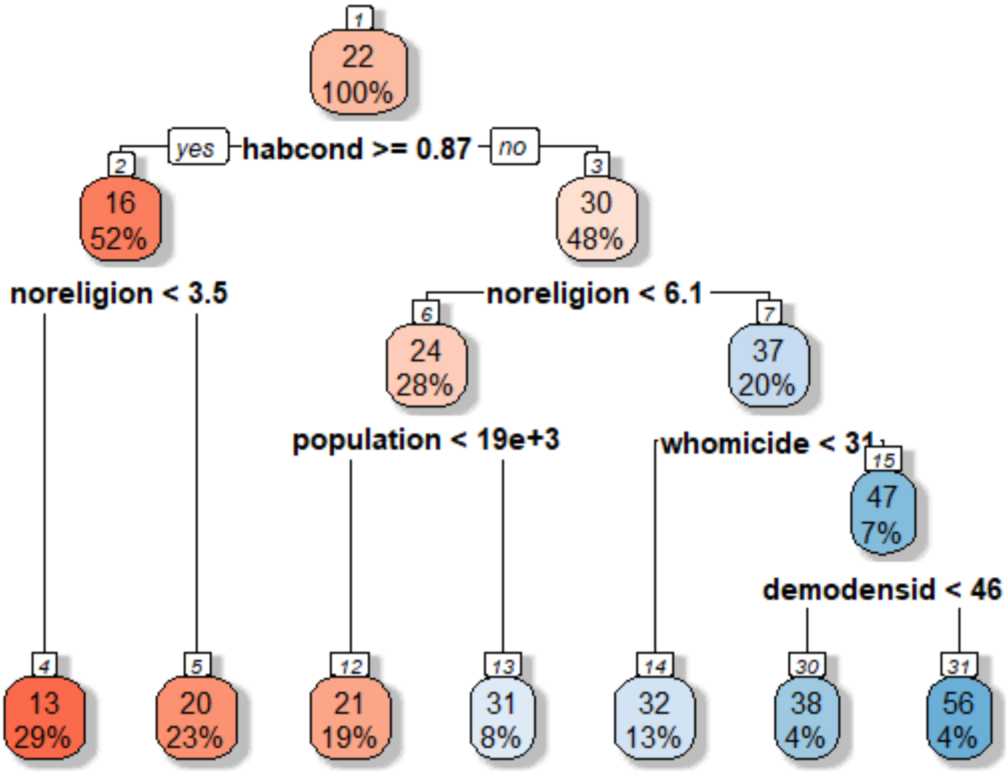


Figure 2: Final tree
Source: elaborated by the authors

The interpretation of the tree is intuitive, the values at the end nodes represent the average of the dependent variable in that group. Thus, the lowest homicide rate is in the area where housing conditions are higher than 0.87, and the percentage of the people with no religion is smaller than 3.5%.

The habitational conditions index (*habcond*) varies between zero and one, being that one represents less housing precariousness and a small number of people living in the area of the house. The final tree points that better housing conditions are linked to a small homicide rate (nodes 4 and 5 of the tree).

A small percentage of people with no religion seems to be linked to small crime rates. This result can be interpreted as a greater moral cost of people with a religious belief. This effect of religion on crime was already detected in crime literature by Ferreira, Bastos, and Betarelli Junior (2019); Shikida (2010); Fajnzylber, Lederman, and Loayza (2002).

Another important result is the presence of the spatial lagged homicide rate in the tree. This corroborate with the hypothesis displacement effect of crime in Brazil already identified in the literature (Almeida, Haddad, and Hewings 2005; T. M. Vital 2018; Ferreira, Bastos, and Betarelli Junior 2019; Shikida 2010). Regions with greater homicide rates are coupled up with the highest homicide rates in the final nodes. Cities with higher populations and higher demographic densities rates are associated with high criminality. This result shows that in average big cities exhibit higher crime rates when compared to small cities.

The final tree has seven end nodes, indicating that the homicide rate in Brazil can be classified in seven classes. Figure 3 presents how this classes are distributed along the Brazillian territory. It is possible to see that regions within the same class are more likely to be linked to one another indicating a spatial pattern corroborating with the importance of the spatial lagged homicide rate from the random forest results.

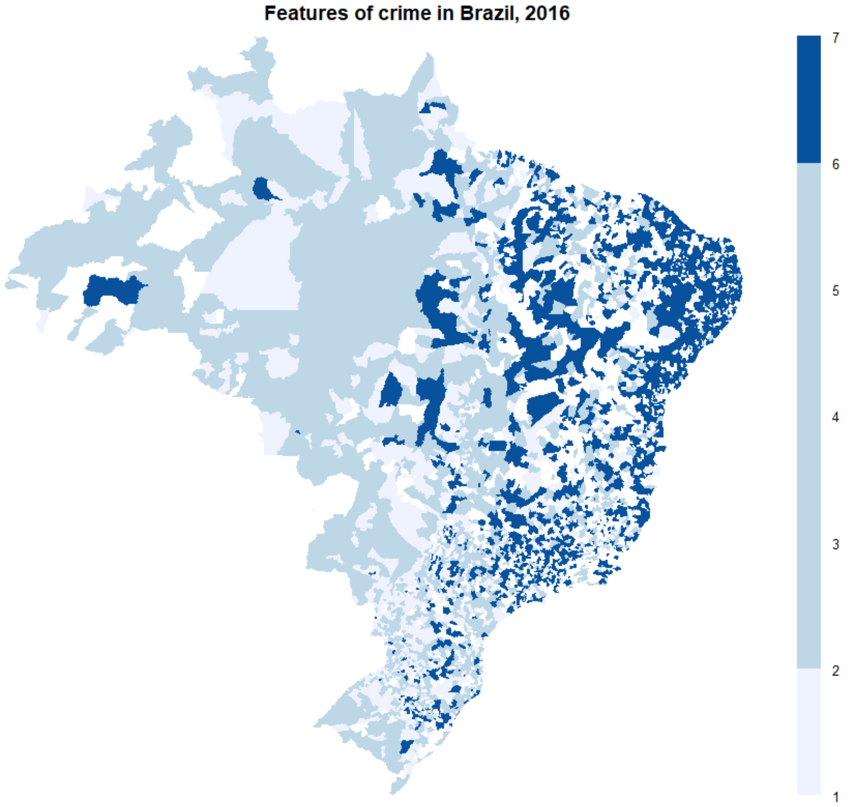


Figure 3: Distribution of the classes along the Brazilian territory
Source: elaborated by the authors.

## 5. Conclusions and implications

In order to determine a causal relationship between public police and the crime rate it is necessary to control for all the variables that can be affecting the latter. However, adding all controls in a single equation can lead to a high dimensionality problem, obfuscating the effect of the treatment. This paper addresses this issue using an alternative approach to classical econometrics.

Using decision tree-based algorithms from statistical learning field, we shed some light in which one the main features that can be linked to homicide rate in Brazil. Housing conditions, population size, demographic density, lack of faith, and the spatial lagged homicide rate are of great importance in determining the 2016 homicide rate in Brazilian municipalities.

The absence of variables related to policing as important to characterize crime, points in a direction that public policies that improve the population's socioeconomic conditions appear to have more effect in fighting crime when compared to an increase in public expenditure of police activity. The spatial lagged homicide rate importance showed that the spatial effect of crime rate in Brazil should be taken into account when evaluating crime.

There is still a long way to go when applying new statistical approaches in economics, this study fits into this context presenting new approaches to problems already studied intensively with other methodologies.

## References


Almeida, Eduardo Simões de, Eduardo Amaral Haddad, and Geoffrey John Dennis Hewings. 2005. "The Spatial Pattern of Crime in Minas Gerais: An Exploratory Analysis." *Economia Aplicada* 9 (1): 39–55.

Alpaydin, Ethem. 2009. *Introduction to Machine Learning*. MIT press.

Athey, Susan. 2015. "Machine Learning and Causal Inference for Policy Evaluation." In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 5–6. ACM.

———. 2018. "The Impact of Machine Learning on Economics." In *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press.

Athey, Susan, and Guido W Imbens. 2019. "Machine Learning Methods That Economists Should Know About." *Annual Review of Economics* 11.

Baumont, Catherine. 2009. "Spatial Effects of Urban Public Policies on Housing Values." *Papers in Regional Science*. https://doi.org/10.1111/j.1435-5957.2008.00200.x.

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." In *The Economic Dimensions of Crime*, 13–68. Springer.

Braga, Anthony A, Andrew V Papachristos, and David M Hureau. 2014. "The Effects of Hot Spots Policing on Crime: An Updated Systematic Review and Meta-Analysis." *Justice Quarterly* 31 (4): 633–63.

Burkov, Andriy, and Michel Lutz. 2019. "The Hundred-Page Machine Learning Book En Français."

Cabral, Maria Viviana de Freitas. 2016. "Avaliação Do Impacto Do INFOCRIM Sobre as Taxas de Homicídios Dos Municípios Paulistas: Uma Aplicação Do Método de Diferenças Em Diferenças Espacial."

Carvalho, José Raimundo, and Sylvia Cristina Lavor. 2008. "Repeat Property Criminal Victimization and Income Inequality in Brazil." *Revista EconomiA* 9 (4): 87–110.

Castro, Talita Egevardt de. n.d. "Políticas de Segurança Pública No Brasil Sob o Olhar Da Economia Do Crime: Os Casos Do PRONASCI e Programa Escola Da Família." Universidade de São Paulo.

Cerqueira, Daniel Ricardo de Castro. 2014. *Causas e Consequências Do Crime No Brasil*. Banco Nacional de Desenvolvimento Econômico e Social.

Chioda, Laura, João M P De Mello, and Rodrigo R Soares. 2016. "Spillovers from Conditional Cash Transfer Programs: Bolsa Família and Crime in Urban Brazil." *Economics of Education Review* 54: 306–20.

Daniel, Cerqueira, Renato Sergio De Lima, Samira Bueno, Luis Iván, Valencia Olaya, Hanashiro Pedro, and Henrique G Machado. 2017. "Atlas Da Violência 2017 Ipea e FBSP." *Ipea e FBSP (Fórum Brasileiro de Segurança Pública)*.

Dias, Camila Caldeira Nunes. 2013. *PCC: Hegemonia Nas Prisões e Monopólio Da Violência*. Editora Saraiva.

Ettensperger, Felix. 2019. "Comparing Supervised Learning Algorithms and Artificial Neural Networks for Conflict Prediction: Performance and Applicability of Deep Learning in the

Field." *Quality and Quantity*. https://doi.org/10.1007/s11135-019-00882-w.

Fajnzylber, Pablo, and Ari Araujo Jr. 2001. "Violência e Criminalidade." *Microeconomia e Sociedade No Brasil*, 333–94.

Fajnzylber, Pablo, Daniel Lederman, and Norman Loayza. 2002. "What Causes Violent Crime?" *European Economic Review*. https://doi.org/10.1016/S0014-2921(01)00096-4.

Ferreira, Sandro de Freitas, Suzana Quinet de Andrade Bastos, and Admir Antonio Betarelli Junior. 2019. "The Role of Social Control in Brazilian Homicide Rates." *Quality and Quantity*. https://doi.org/10.1007/s11135-019-00886-6.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2001. *The Elements of Statistical Learning*. Vol. 1. Springer series in statistics New York.

Furqan, Maham, and Haider Mahmood. 2020. "Does Education Reduce Homicide? A Panel Data Analysis of Asian Region." *Quality & Quantity*. https://doi.org/10.1007/s11135-020-00981-z.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. Vol. 112. Springer.

Kelly, Morgan. 2000. "Inequality and Crime." *Review of Economics and Statistics* 82 (4): 530–39.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. 2015. "Prediction Policy Problems." *American Economic Review* 105 (5): 491–95.

Lantz, Brett. 2015. *Machine Learning with R*. Packt Publishing Ltd.

Levitt, Steven D. 1997. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime." *American Economic Review*. https://doi.org/10.1257/00028280260344777.

———. 2004. "Understanding Why Crime Fell in the 1990s: Four Factors That Explain the Decline and Six That Do Not." *Journal of Economic Perspectives*. https://doi.org/10.1257/089533004773563485.

———. 2017. "The Economics of Crime." *Journal of Political Economy*. https://doi.org/10.1086/694630.

Manso, Bruno Paes, and Camila Nunes Dias. 2018. *A Guerra: A Ascensão Do PCC e o Mundo Do Crime No Brasil*. Editora Todavia SA.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31 (2): 87–106.

Murray, Joseph, Daniel Ricardo de Castro Cerqueira, and Tulio Kahn. 2013. "Crime and Violence in Brazil: Systematic Review of Time Trends, Prevalence Rates and Risk Factors." *Aggression and Violent Behavior* 18 (5): 471–83.

Pereira Filho, Oliveira Alves, and Maria Da Conceição Sampaio De Sousa. 2018. "Avaliação De Impacto Das Guardas Municipais Com O Uso De Tratamentos Binários, Multivalorados E Contínuos." In *Anais Do XLIV Encontro Nacional de Economia [Proceedings of the 44th Brazilian Economics Meeting]*. ANPEC-AssociaÃ § Ã£ o Nacional dos Centros de PÃ³s-GraduaÃ § Ã£ o em ….

Shikida, P.F.A. 2010. “Considerações Sobre a Economia Do Crime No Brasil: Um Sumário de 10 Anos de Pesquisa.” *Economic Analysis of Law Review*. https://doi.org/10.18836/2178-0587/ealr.v1n2p318-336.

Torgo, Luis. 2003. “Data Mining with R.” *Learning by Case Studies. University of Porto, LIACC-FEP. URL: Http://Www. Liacc. up. Pt/Ltorgo/DataMiningWithR/. Accessed On* 7 (09).

Vital, Tauã Magalhães. 2018. “Uma Análise de Impacto Da Guarda Municipal No Brasil.”

Vital, Tauã, Daniel Morais De Souza, and Jessica Faciroli. 2020. “Unemployment, Poverty and Police Performance: An Ardl Analysis of Crime in Sao Paulo.” *Economics Bulletin*.

Wilson, James Q, Joan Petersilia, and James Q Wilson. 2002. *Crime: Public Policies for Crime Control*. Ics Press San Francisco, CA.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT press.