Identifying Correlates of Homicide Rates in Michoacán, Mexico

Carl M. Kruger(✉) and Matthew S. Gerber

Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA, USA
{cmk3pt,msg8u}@virginia.edu

Abstract. Violent crime rates in Mexico have reached epidemic levels recently due to a corrupt and ineffective government and violent drug-trafficking organizations (DTOs). This research analyzes social, economic, and demographic data, to identify factors that correlate with homicide rates. Previous research efforts have focused on trends of homicide rates and provided subjective recommendations without supporting empirical evidence. This research provides an objective analysis of homicide rates, using a large database consisting of census data provided by the Instituto Nacional De Estradstica Y Geografa (INEGI) website. Through formal experimentation, we found that the most significant indicators were economic and education variables (e.g., Occupants in Refuge and Population over 5 Years Without Schooling), supporting recommendations on policy changes from other research efforts. This paper presents our experiments, results, and insights regarding future work.

Keywords: Homicide · Mexico · Linear regression

1 Background

The United Nations Office on Drugs and Crime ranks Mexico as one of the most dangerous countries in the world, as the twenty-second country with the highest homicide rate [4]. A major source of violence in Mexico stems from the drug war. Confluence of rival Drug Trafficking Organizations (DTOs) and government corruption has increased violent crime rates to epidemic proportions in Mexico. In December 2006, President Felipe Calderon mobilized the Mexican military to engage in a major offensive with DTOs in the state of Michoacán. The drug war continues today and has claimed well over 50,000 lives [1].

Previous research efforts offer expert analysis of the history of the Mexican Drug War as well as recommended solutions to reverse the trends of violence in Mexico. The Justice in Mexico Project (JMP) is an organization that addresses security and human rights issues related to drug violence.[1] They draw few conclusions and recommendations based on data analysis but rather make a subjective analysis of the increase in drug-related homicides as a result of corrupt

[1] https://justiceinmexico.org/

N. Agarwal et al. (Eds.): SBP 2015, LNCS 9021, pp. 321–326, 2015.
DOI: 10.1007/978-3-319-16268-3_37

Mexican politics [1]. Likewise, Sibel McGee, et al., discuss the Mexican Cartel problem from a systemic perspective [3], breaking down the problem into several domains containing dynamic relationships and critical variables involved in DTO operations. They provide a holistic assessment of the cartel problem with recommendations on long-term solutions to combat cartel aggression, to include education and economic reform.

This paper presents research aimed at identifying correlates of homicide rates in Michoacán, Mexico. Using linear regression methods, we analyzed indicator variables that correlate with homicide rates. Our results show there are strong correlations between education, economics, and homicide rates. Our results substantiate other research efforts and current plans by the Mexican Government to reduce homicide rates as well as highlight other areas not previously explored. To date, such plans have not been backed by substantial amounts of empirical research.

2 Problem and Objectives

We have focused our research on identifying and understanding correlates of homicide rates in Mexico, focusing on the state of Michoacán. Michoacán continues to be relevant in the drug war, as the Mexican military as well as emerging vigilante groups continue to engage DTOs [5] [6]. Our primary objective was to empirically identify correlates of homicide rates within Michoacán, and highlight areas not previously addressed. By identifying such correlates and quantifying their strengths, we believe decision makers will be in a better position to understand the recent increase in homicides within Michoacán and develop effective solutions. Although the drug war is one of the primary motivating factors, this research considers homicides of all types.

3 Data and Methodology

We obtained data from the Instituto Nacional De Estradstica Y Geografa (INEGI) website.[2] INEGI publishes data, consisting of 965 indicators in four categories: (1) Economy, (2) Environment, (3) Population, Households, and Housing, and (4) Society and Government. We aggregated the data at the municipality level and used homicide rate per 100,000 inhabitants as our response variable. All other variables were also normalized to reflect a rate per 100,000 inhabitants. INEGI measures most indicators annually, but some are measured in increments of five years. For the latter, years with no measurement assume the value of the most recent measurement. We re-normalized the data on a scale of 0 to 1, where 0 corresponds to the minimum value for each indicator and 1 corresponds to the maximum value for each indicator. Other crime rates strongly correlated with homicide rates; however, since our intent is to identify social correlates of homicide, we removed these variables. Additionally, the economic variables oriented

[2] http://www.inegi.org.mx

around publishing statistics of the various economic industries in Michoacán, Mexico, and did not provide statistics of individual economic strength or poverty levels, so we removed these variables as well. Any indicators with missing values were removed from the data set, yielding 149 indicators for analysis.

Focusing on 2000-2010, there are only 11 observations per municipality. We pooled observations from all municipalities, resulting in a data set with 1,243 observations. We fit a linear regression model to the response and indicator variables using the generic form of linear regression and performed two analytic tasks based on the resulting model. First, we assessed the indicator variables for significance and evaluated their weights in the context of prior research findings. Second, we evaluated the linear regression model in a 10-fold cross-validation experimental setup designed to assess the model's ability to explain previously unseen data.

4 Results and Discussion

We developed a main effects model using all 149 variables, and then we conducted stepwise regression. The model showed evidence of multicollinearity, so we removed them in a stepwise manner with a variance inflation factor cutoff of 5 [10]. The resulting model consisted of 38 variables and was statistically significant at a level of 0.05. Additionally, we calculated both the R^2 and adjusted R^2 values to determine each model's accuracy. The R^2 was 0.458 and Adjusted R^2 was 0.441. The diagnostic plots in Figure 1 show some evidence of heteroscedasticity and non-normality in our data. This is also evident in the Residuals vs. Leverage Plot as the influential points in the data consist of extremely high homicide rate observations.

We calculated the p-value for each indicator variable using the same statistical significance threshold, and 15 of the 38 variables were statistically significant. Table 1 shows a summary of the most interesting, statistically significant variables, their coefficients, and statistical significance. The population of men and general deaths of women being positively correlated with homicide rates could suggest that a larger male population accounts for higher homicide rates. Economically, with occupants in refuge positively correlated with homicide rates, it might suggest increased economic hardships force more families into refuge, increasing vulnerability to criminal activity. Both education variables show how a less educated population positively correlates with homicide rates, while a more educated population negatively correlates with homicide rates. Finally, Private Homes with 5-8 Inhabitants could suggest that these are extended family homes. Being negatively correlated with homicide rates, extended family homes could be less likely to fall prey to criminal activities due to strong family structure.

We used 10-fold cross-validation to evaluate how well the model performed on different subsets of data, using Mean Squared Error (MSE) to evaluate our model's performance. The MSE was 452. In order to investigate what might account for such high MSEs, we broke down the results by municipality. The five municipalities with the highest MSEs recorded some of the highest homicide rates. The 95% confidence interval of all homicide rate observations falls

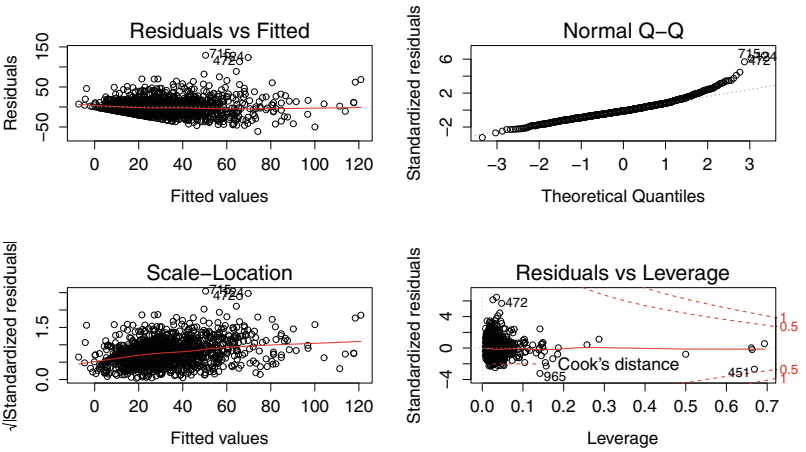

Fig. 1. Diagnostic Plots for Linear Regression Model

Table 1. Variable Statistics

Variable	Coefficient	Stat. Sig.
Total Population, Men	61.470	< 0.001
General Deaths, Women	15.574	0.0159
Occupants in Refuge	10.690	< 0.0182
Private Homes With 5-8 Inhabitants	-30.525	< 0.001
Population Over 5 Years Without Schooling	22.061	< 0.001
Students in Basic and Higher Secondary Education	-19.208	0.0131

between 0 and 100. Over 30% of all observations of the five municipalities with the highest MSE results recorded homicide rates above 100. Meanwhile, the five municipalities with the lowest MSE results remained well within this 95% confidence interval, (Table 2). This suggests that the model does not perform well in cases of extremely high homicide rates. In cases with more moderate homicide rates the model performs better.

Our research corroborates research done by McGee, et al., which posits that economic and education reform are paramount in developing a population that will resist the allure of criminal business [3]. However, neither McGee, et al., nor JMP address the male population as a primary concern for criminal behavior. Based on prison statistics for Michoacán, Mexico, over 85% of inmates are male.[3] This suggests criminal activity is predominantly executed by male citizens. Additionally, another aspect not previously addressed is the impact of

[3] These statistics provided by INEGI Bank of Information for 2009-2011.

Table 2. MSE and Homicide Rate Statistics

Municipality	MSE	Min H.R.	Mean H.R.	Max H.R.	H.R. > 100
La Piedad	54.5	10.6	31.2	49.4	0
Contepec	70.3	10.0	28.4	52.1	0
Tzintzuntzan	77.1	0.0	20.4	40.8	0
Ocampo	81.0	10.6	23.8	37.2	0
Zamora	82.7	3.7	28.6	49.2	0
Tumbiscato	1279.6	0.0	36.3	131.5	1
Aquila	1624.7	13.5	71.4	153.0	2
Taretan	1705.1	22.6	67.8	179.0	2
Nocupetaro	1795.4	13.1	85.0	128.0	5
Coahuayana	2860.6	21.5	112.5	193.2	7

extended family homes on criminal activity. This could suggest that areas with more extended family households develop stronger family networks that are more immune to the allure of DTOs.

5 Conclusions and Future Work

Violence in Mexico continues to be a primary concern for the Mexican and United States governments. Policy makers and humanitarian efforts have called for reform in various areas (e.g., education), but there has been a lack of empirical research supporting such reforms. We have developed a linear model of homicides within Michoacán based on data collected by INEGI. Our model provides empirical support for reform in areas like education and economics, corroborating past research [1] [3]. Our model also highlights other areas of reform that have not received as much attention (e.g. Total Male Population and Extended Family Homes). Efforts to address some of these factors are already under way. In February 2014, the Peña Nieto administration launched a social development program called "Plan Michoacán," focused on key areas of economic development, education, infrastructure and housing, public health, and social development and sustainability [2]. The results of this study substantiate their areas of focus and possibly identify other areas that were not included in Plan Michoacán.

Future work will focus on three areas. First, we will investigate interactions between independent variables. Second, we will explore alternative modeling techniques including Random Forest Regression and its associated variable importance measures [9]. While linear regression offered interesting results, MSE was high for some municipalities, which may be a result of non-linearities. Random Forest Regression may overcome this shortcoming and offer greater insight as to correlating factors. Third, we hope to focus more specifically on cartel-related homicides. The present investigation focused on homicides generally because such data are readily available. However, a large component of the homicide process is driven by cartels. By focusing specifically on cartel-related homicides, we hope to uncover

key contributory factors that are not currently being considered by anti-violence efforts in Mexico.

References

1. Shirk, D.A., et al.: Drug Violence in Mexico: Data and Analysis from 2001–2009. Technical Report, Justice in Mexico Project, Trans-Border Institute, Joan B. Kroc School of Peace Studies (2010)
2. Heinle, K., et al.: Drug Violence in Mexico: Data and Analysis through 2013. Technical Report, Justice in Mexico Project, Trans-Border Institute, Joan B. Kroc School of Peace Studies (2014)
3. McGee, S., et al.: Mexico's Cartel Problem: A Systems Thinking Perspective. Technical Report, Applied Systems Thinking Institute, Analytic Services, Inc. (2014)
4. Global Study on Homicide, United Nations Office on Drugs and Crime. http://www.unodc.org.gsh
5. The War on Michoacán: A Brief Chronology. http://www.borderlandbeat.com/2014/01/the-war-in-michoacan-brief-chronology.html
6. Flannery, N.P.: Mexico Media Roundup: Another Month of Violence and Vigilantes in Michoacán. http://www.forbes.com/sites/nathanielparishflannery/2014/03/31/
7. Maindonald, J.H., et al.: DAAG: Data Analysis and Graphics data and functions. R package version 1.20 (2014). http://CRAN.R-project.org/package==DAAG
8. Alfons, A.: cvTools: Cross-validation tools for regression model. R package version 0.3.2 (2012). http://CRAN.R-project.org/package==cvTools
9. Liaw, A., et al.: Classification and Regression by randomForest. R News **2**(3), 18–22 (2002)
10. Fox, J., Weisberg, S.: An R Companion to Applied Regression, 2nd edn. Sage, Thousand Oaks. http://socserv.socsci.mcmaster.ca/jfox/Books/Companion