An Economic Analysis of Crime on Student Performance

By:

Justin Yu - 1006747111 - justi.yu@mail.utoronto.ca

Hamza Khan - 1005908886 - hamz.khan@mail.utoronto.ca

Eco 375: Applied Econometrics

University of Toronto

Department of Economics

Draft of Assignment 1

Tuesday April 4th, 2023

Abstract

The issue of violence and its impact on academic performance deserves our attention and consideration. Educators and community members must prioritize the well-being of students and create a supportive and nurturing environment where they can flourish academically and personally. This study aims to examine the relationship between crime and academic success using regression analysis. The study utilizes a sample of publicly funded Chicago schools across various communities, with a number of crimes and academic achievement on the Math Illinois Standard Achievement Test (ISAT) serving as the key variables of interest. The goal was to determine whether the two variables have a statistically and economically meaningful relationship. Although the coefficients for academic accomplishment were statistically significant, it was deemed not economically significant. Therefore, the regression analysis's findings conclude that there is no meaningful relationship between crime and academic achievement. The implications of these findings may help policymakers and educators understand the value of addressing underlying socio-economic problems that may obstruct academic success rather than limiting their attention to lowering criminal activity as a solution.

Introduction:

Communities all over the globe struggle with the enduring and pervasive problem of crime. The social and economic fabric of the community as a whole, as well as the safety and well-being of residents, may all be significantly impacted. One group of the population that should be of concern is students and their education in an area where the effects of crime are especially noticeable.

Many studies have examined the relationship between community violence, academic performance, and school accountability. In one study, researchers examined information from the Chicago Public Schools district for the years 2008 to 2015. The study concludes that student academic success is negatively impacted by exposure to community violence, especially in schools with low academic performance (Casey, Schiman, & Wachala, 2018). Another study conducted by O'Brien et al (2021), demonstrated that criminal activity can harm educational outcomes, such as academic performance, graduation rates, and total educational attainment. We know that student mental health is an important factor behind academic performance and that crimes of violent nature can have a negative psychological impact on a student's mental health, which is then detrimental to their performance at school (Assari et al, 2018). Furthermore, Boxer et al (2020) conducted a geospatial analysis on the effects of neighborhood violent crime and student performance where they concluded that interventions targeting both neighborhood safety and school-level resources and support are necessary. Despite the fact that the connection between crime and education is becoming more widely acknowledged, (given the recency of these previous studies), a further, more in-depth study is still required. This paper seeks to address this gap by exploring the relationship between crimes and their effects on student performances (i.e. their academic achievements) in exams.

To determine how much crime impacts academic achievement and other measures of educational success, we will specifically analyze data on crime rates and educational outcomes in a number of communities. We will also take into account other potential controls, such as socioeconomic position. With the help of our analysis, we hope to shed new light on the intricate relationship between crime and education and pinpoint possible solutions to the problems that crime poses in educational settings. We believe this research has important implications for policymakers, educators, and other stakeholders working to promote safer and more prosperous communities.

The Context and Data:

The data used in this paper is panel data spanning from the years 2005 to 2013. The data includes the overall academic performance of students in grades 3-8 writing the math portion of the Illinois Standard Achievement Test (ISAT) from 415 different publicly funded Chicago schools and 76 communities. The academic performance of a school is recorded in three categories as follows; “Above Expectation”, “Meeting Expectation” and “Warning” and the statistics are presented as the proportion of students in each of the respective categories. In addition, the panel data includes metrics for the number of violent/property crimes committed in a 2-week period before, during, and after the day the exam was written. The number of violent crimes is expanded from a 0.1-mile radius to a 0.5-mile radius from the location of the given Chicago public school. The data also includes control/fixed variables such as the median income of the households, proportion of whites, median age, median household size, the proportion of people who identify as Hispanic, and the proportion of the people below poverty in a given community during a chosen year.

The main variable of interest is the number of violent crimes committed 2 weeks before the ISAT was held. Variation in our explanatory variable comes from the observational data itself meaning that it is a result of the natural differences and fluctuations that exist in the data set. The variation also changes depending on how far (radius) from the given school it was recorded. From Table 1, we can see the spread of violent crimes within 0.1 miles of a school is around 0.913 standard deviations of the average 0.537 crimes near 0.1 miles of a Chicago school.

For the analysis, our population of interest is the publicly funded schools in Chicago between the years 2005 to 2013. The outcome of interest is the proportion of students in the “warning” category for the math portion of the ISAT. Table 1 shows a summary of statistics for the panel data with 3662 total observations and 415 unique schools. According to the central limit theorem, in large samples, the sampling distribution of $\bar{Y}$ is approximately normal. Due to our large sample size, we are able to justify the use of certain statistical tests that need to assume normally distributed errors, such as hypothesis and confidence interval tests. Some interesting points to note from Table 1 are that within Chicago, there seem to be communities that have an unemployment rate of 1, or rather 100% of the people are unemployed in that community. Also, we can see that there are communities where 100% of the residents are below the poverty line, there are predominantly White communities (98.1%) as well as predominantly Hispanic (98.4%).

Regression analysis:

Simple Regression Model:

We begin our analysis by estimating a simple regression model between the proportion of students which were placed in the "warning" category in a given school (Y_i) and the number of violent crimes within 0.1 miles of each school (X_i). Our baseline model is a simple linear regression (SLR) with one predictor; the number of violent crimes committed in a 0.1-mile radius in the period 2 weeks before the test was held (Table 2, column (1)). This regression can be modeled by the equation: $Y_i = \beta_0 + \beta_1 X_i + u_i$

From this simple regression model, we obtained an estimated coefficient, $\hat{\beta}_1$ to be 0.0377 with a standard error of 0.002 meaning that a unit increase in a count of violent crime being committed near 0.1 miles of a school two weeks before the Math ISAT increases the proportion of students which obtained in the warning performance category by 3.8%. The relationship seems to be relatively small, but using hypothesis testing, we obtained a t-value of 19.84 which indicates a high level of statistical significance (i.e. $\beta \neq 0$). The slope coefficient is significant statistically, however, it is relatively small in terms of economic significance. Take the standard deviation of X_i to be 0.913, (Table 1) which means that a one standard deviation change in X_i would be a difference of 0.0344 in Y_i. Meanwhile, the standard deviation in Y_i is found to be 0.11 (Table 1). This means that a difference of 1 standard deviation of X_i, only explains 31% of the standard deviation in Y_i. Since the standard deviation of Y_i is already small, the economic significance of X_i is even smaller.

To properly use hypothesis testing and confidence intervals, the assumptions of normality and homoscedasticity (equal variance of residuals across all levels of predictors) should hold. Homoscedastic errors are a strong assumption that is not really seen using real-world data such as ours. The regression model column (1) assumes homoscedasticity, but we know that this is probably not true since the model does not account for any within-school correlations and heteroscedasticity of the errors, due to panel data. To account for the violation in homoscedasticity, column (2) clusters our data around schools and from this, a higher standard error of 0.00257 was obtained but it is still a significant coefficient at the 1% level using hypothesis testing. Using clustered standard errors only reduces some model misspecifications.

Extending the Model:

One of these model misspecifications is that columns (1) and (2) only look at the effects that violent crime can have on students' performance in the Math ISAT. But what if other variables affect Y_i and X_i that are not included in the model? For example, think about a school's quality. The quality of a school such as its teaching staff, culture, and learning resources all have an impact on student performance in the ISAT. The quality of the school can also be linked to the number of violent crimes being committed near a school as schools of "higher quality" have better resources to deal with crime. Without including differences in school quality, the effect of X_i on Y_i may be inaccurately estimated as the effect may be confounded by the differences in school quality. This affects the least squares assumption 1, which states that the predictor X_i is uncorrelated with the error term u_i. Not including school quality in the model would mean that X_i would be correlated with u_i and thus $E[u_i|X_i] \neq 0$ and $E[\hat{\beta}] \neq \beta$. This just means that school quality should be added to the model, however, observing/measuring school quality is difficult.

This is where fixed effects come in as they help control for these unobservable outcomes that may be correlated with the explanatory or outcome variables. By using school-fixed effects in a regression model, we can account for unobserved school-level factors that are assumed to be constant over time and may bias our estimates of the effect of other variables. Table 2 column (3) looks at a model which includes school-fixed effects, from this, we can see that with the addition of school-fixed effects, there is a decrease in the coefficient $\hat{\beta}_1$. This decrease from 0.038 to 0.012 indicates a previous upward omitted variable bias which is now controlled for. The within R^2 obtained was smaller, however, the coefficient still remains statistically significant at the 1% level with a smaller standard error. Time-fixed effects can also be used to control for unobserved time-invariant factors which can affect the outcome variable. For example, any changes in school policies within the years (2005 - 2013) which would have an impact on Y_i are accounted for. Realistically, school quality can change year to year as teachers are replaced, curriculum changes, and schools expand. Table 2 column (4) reflects this addition of a time trend however, there seems to be no effect on the model indicating that the explanatory variables are not affected by the time-invariant factor.

Now that we have controlled for unobservable variables by comparing different schools across years, we can control for other measured variables which impact the regression model. Socioeconomic variables like the average household size, the unemployment rate, per capita crime rate can affect the proportion of students in the performance category of "Warning" and the count of crimes being committed near schools. Another important variable to control for is the level of crime being committed during and after the testing period. Perhaps the level of crime changes during and after the testing period as police may patrol more during the period or students have more free time to commit crimes after the period. Controlling for these variables isolates the true effect violent crime has on student performance and in Table 2 column (5), we see a smaller but still significant $\hat{\beta}_1$ of 0.0091. So adding in all these controls seems to have increased the accuracy of our estimate since the standard error is now lowered. A within R^2 of 0.059 was obtained which indicates that the independent variables in the fixed effects model only account for 5.9% of the variation in the dependent variable. However, there seem to be no coefficients for some of the control variables included in model (5). This may be because the control variables are collinear with the absorbed fixed effects. Looking back at the dataset we do in fact see that these control variables were constant across time for different schools and this effect is accounted for by using the fixed effects model. This presents itself as a problem since we cannot force the control variables into the model unless we abandon school and time-fixed effects. So for our interest, we would need to drop those controls from our fixed effect model.

But what happens if we extend the region at which violent crimes are being committed near a Chicago school? If we were to analyze the effect as the distance from the school increases we would need to categorize crime count to minimize a violation in the multicollinearity least square assumption. This is because the count of violent crimes occurring within 0.2 miles of a school includes crimes that occurred within 0.1 miles of a school. We can see this relationship between the variables more accurately using the correlation matrix (Table 3). To reduce multicollinearity between the variables, we decided to split the count of violent crime into different categories of range from a school. Since the variables are now less likely to be correlated, we have a model of: $Y_{it} = \beta_1 X_{1,it} + \beta_2 X_{2,it} + \beta_3 X_{3,it} + \beta_4 X_{4,it} + \eta_{it} + a_i + u_{it}$

Variable Y_{it} represents the proportion of students which were placed in the warning performance category. The variable $X_{1,it}$ is the count of violent crimes being committed near 0.1 miles of a school for school id, i = 1, …, n in time t = 2005, …, 2013. The variable $X_{4,it}$ is the count of violent crimes being committed near 0.3- 0.5 miles of a school for school id, 'i'. The variable η_{it} is the remaining control variables for a school id,i in time, t. Variable a_i are school specific intercepts and $u_{it}$is the error associated with school id, i in time, t. Table 4 column (1) shows us the results of running this regression on distance categories of violent crimes near to a school. From the model, we do not see a decreasing relationship between violent crimes being committed near a school and student performance in the Math ISAT as there is a stronger effect for violent crimes committed within the range of 0.3-0.5 miles of a school than there is for crimes committed from 0.0-0.3 miles. From the slope coefficient of violent crimes being committed near 0.1 miles of a school, we can see that for an increase in violent crime two weeks before an exam, there is a statistically significant (at the 10% level) increase of 0.28 percent of students who get a performance academic warning. This effect decreases but then increases in the range of 0.2-0.3 miles from a school. This can be explained as students usually don't live between 0.0-0.2 miles from their school. So, although the effect of violent crimes in those areas can affect students due to its proximity to the school, since most students do not live within those ranges, the effect is lessened. However, the further out from school, the effect seems to increase, this is probably because that is where we start to see the residential areas of some students and they are affected by any violent crimes happening near their neighborhoods.

We can also regress Y_{it} on property crimes shown in Table 4 column (2). From this regression, we do see that property crimes have a smaller effect on Y_{it} than violent crimes. The relationship, however, does not seem to be significant near schools and decreases the further away from 0.1 miles to the school. Since property crimes include crimes where people were not harmed, we do not expect a larger or more significant effect on Y_{it} as opposed to violent crimes, which have a stronger psychological effect on students as discussed in Assari et al (2018). Including both violent and property crimes into a regression model, we found similar results from before, with violent crimes being a stronger and more significant predictor of student performance in the Math ISAT than property crime.

Limitations of Results:

The analysis presented in this study is subject to several internal and external validity issues that should be taken into consideration when interpreting the findings.

Externally: While the Math ISAT is a widely used test, it does not fully capture academic achievement making it hard to generalize the findings in this study for other tests. Also, the study only included Chicago data making it not applicable to other nations or even cities in the US.

Internally: We removed socioeconomic factors from the model as they were constant across time. However, we know that socioeconomic factors do not actually remain constant over time unlike the data given by the provider, so removing these variables from the model means that there still lies an omitted variable bias problem. Adding in more variables to try and control for this, however, increases the possibility of imperfect multicollinearity which affects the estimation of the regression coefficients. Lastly, our OLS assumptions were discussed and some adjustments were made to reduce potential violations, but these assumptions were not fully tested if held. Formal checks for outliers in independent and dependent variables along with normality, homoscedasticity, and multicollinearity tests can help validate these assumptions.

Conclusion:

This study used regression analysis to look into the connection between community violence and academic achievement. We found that violent crimes committed near 0.1 miles of school can have a positive, statistically significant impact on the proportion of students which got a performance warning in the Math ISAT. Property crime had a lesser but still significant effect, although not at distances less than 0.1 miles. Both effects, however, were deemed not economically significant and got smaller with the addition of control variables and fixed effects. We did, however, observe varying effects of crime at different distances from school. This result agrees with earlier studies conducted by O'Brien et al (2021) that link community violence to academic achievement, but contrasts with other studies that showed crime having a lesser impact on performance, further away from school (Casey et al (2018)). Despite some limitations, the findings of this study demonstrate the need for more investigation into the connection between violence and academic achievement. The ultimate objective of this study was to raise academic achievement awareness for students, who live in crime-ridden communities. Future research could expand on this analysis' limitations by looking at a broader variety of factors or statistical methods that might affect students' academic performance.

References:

Assari, S., & Moghani Lankarani, M. (2018). Violence Exposure and Mental Health of College Students in the United States. *Behavioral sciences (Basel, Switzerland)*, *8*(6), 53. https://doi.org/10.3390/bs8060053

Boxer, P., Drawve, G., & Caplan, J. M. (2020). Neighborhood violent crime and academic performance: A geospatial analysis. American Journal of Community Psychology, 65(3-4), 343–352. https://doi.org/10.1002/ajcp.12417

Casey, M, Schiman J. C., and Wachala M. 2018. "Local Violence, Academic Performance, and School Accountability." AEA Papers and Proceedings, 108: 213-16. DOI: 10.1257/pandp.20181109.

Milam, A. J., Furr-Holden, C. D., & Leaf, P. J. (2010). Perceived school and neighborhood safety, neighborhood violence and academic achievement in Urban School Children. The Urban Review, 42(5), 458–467. https://doi.org/10.1007/s11256-010-0165-7

O'Brien, D. T., Hill, N. E., & Contreras, M. (2021). Community violence and academic achievement: High-crime neighborhoods, Hotspot Streets, and the geographic scale of "community." PLOS ONE, 16(11). https://doi.org/10.1371/journal.pone.0258577

Stock, J. H., & Watson, M. W. (2020). Introduction to econometrics. Pearson.

Tables and Figures:

Table 1:

Summary Statistics of Variables Used in Analysis

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
Prop Students in Warning Category	3,662	0.142	0.110	0	0.671
Property Crimes within 0.1 miles	3,662	0.531	0.853	0	7
Property Crimes within 0.2 miles	3,662	2.431	2.323	0	23
Property Crimes within 0.3 miles	3,662	5.517	4.412	0	45
Property Crimes within 0.5 miles	3,662	14.59	9.703	0	103
Violent Crimes within 0.1 miles	3,662	0.537	0.913	0	7
Violent Crimes within 0.2 miles	3,662	1.954	2.149	0	18
Violent Crimes within 0.3 miles	3,662	4.167	3.754	0	34
Violent Crimes within 0.5 miles	3,662	10.87	8.403	0	69
Average Household Size	3,662	2.901	0.644	1.340	4.550
Median Age Population	3,662	33.99	6.123	15.40	49.90
Median Household Income	3,662	45,080	21,901	5,725	125,100
Proportion Below Poverty Line	3,662	0.218	0.148	0	1
% Hispanic	3,662	0.275	0.315	0	0.981
% White	3,662	0.391	0.334	0	0.984
Unemployment Rate	3,662	0.175	0.103	0.0215	1
Average Per Capita Crime Rate	3,662	0.0354	0.0308	0.00229	0.268

Note: The unit of observation is school-year for 415 different Chicago schools from the years 2005 to 2013.

Table 2:

Proportion of Students in the Warning Performance Category
Panel Data, Fixed Effects

VARIABLES	(1)	(2)	(3)	(4)	(5)
Violent Crimes within 0.1 miles	0.0377***	0.0377***	0.0116***	0.0116***	0.00910***
	(0.00190)	(0.00257)	(0.00192)	(0.00192)	(0.00186)
Average Household Size = o,					-
Median Age Population = o,					-
Median Household Income = o,					-
Proportion Below Poverty Line = o,					-
Unemployment Rate = o,					-
Average Per Capita Crime Rate = o,					-
% White = o,					-
% Hispanic = o,					-
Violent Crimes during ISAT					0.0131***
					(0.00159)
Violent Crimes after ISAT					0.00880***
					(0.00178)
Constant	0.122***	0.122***	0.136***	0.136***	0.123***
	(0.00201)	(0.00404)	(0.00103)	(0.00103)	(0.00165)
Observations	3,662	3,662	3,662	3,662	3,662
R-squared	0.097	0.097	0.016	0.016	0.059
Clustered Errors	NO	YES	YES	YES	YES
School FE	NO	NO	YES	YES	YES
Year FE	NO	NO	NO	YES	YES
Controls	NO	NO	NO	NO	YES

Note: The regression models in this table shows the building up and extension our simple regression model to a model which includes with multiple regressors and fixed effects to control for omitted variable biases. To account for model heteroscedasticity, standard errors are clustered at the school level.

*** indicates a p-value of less than 0.01 significance level, ** indicates a p-value of less than 0.05 significance level, and * indicates a p-value of less than 0.1 significance level

Table 2:

Matrix of correlations

Variables	(1)	(2)	(3)	(4)
(1) Violent Crimes within 0.1 miles	1.000			
(2) Violent Crimes within 0.2 miles	0.659	1.000		
(3) Violent Crimes within 0.3 miles	0.553	0.831	1.000	
(4) Violent Crimes within 0.5 miles	0.468	0.692	0.859	1.000

Table 4:

OLS Estimation of The Proportion of Students in the Warning Performance Category
Panel Data, Fixed Effects

VARIABLES	(1)	(2)	(3)
Violent Crimes within 0-0.1 miles	0.00276*		0.00261
	(0.00164)		(0.00162)
Violent Crimes within 0.1-0.2 miles	0.00268**		0.00267**
	(0.00109)		(0.00108)
Violent Crimes within 0.2-0.3 miles	0.00353***		0.00349***
	(0.000874)		(0.000863)
Violent Crimes within 0.3-0.5 miles	0.00373***		0.00358***
	(0.000378)		(0.000379)
Property Crimes within 0-0.1 miles		0.000127	0.000274
		(0.00161)	(0.00144)
Property Crimes within 0.1-0.2 miles		0.00271***	0.00131*
		(0.000883)	(0.000792)
Property Crimes within 0.2-0.3 miles		0.00246***	0.00118*
		(0.000693)	(0.000620)
Property Crimes within 0.3-0.5 miles		0.00193***	0.000774**
		(0.000381)	(0.000341)
Constant	0.0365***	0.0798***	0.0217***
	(0.00521)	(0.00569)	(0.00600)
Observations	3,662	3,662	3,662
R-squared	0.246	0.052	0.255
R-squared	0.246	0.0519	0.255
Controls	YES	YES	YES

Note: The count crimes were categorized based on their distance from a school, in order to reduce multicollinearity. Controls are also included in each regression model, and to account for model heteroscedasticity, standard errors are clustered at the school level.

*** indicates a p-value of less than 0.01 significance level, ** indicates a p-value of less than 0.05 significance level, and * indicates a p-value of less than 0.1 significance level