The Correlation Between Educational Attainment and Police Stops: A Logit Model Analysis

Ziqi Lai *

Mathematics with Applied Mathematics, Faculty of Science and Engineering, University of Nottingham Ningbo China, 315100 Ningbo, China

* Corresponding Author Email: smyzl15@nottingham.edu.cn

Abstract. While prior studies on police stops have emphasized visible social characteristics, less attention has been given to the role of educational attainment. This study examines the influence of education on the probability of being stopped by police, using 4,488 individual data in 1989 of the National Longitudinal Survey of Youth. This study employs a logit regression model to assess the probability of police contact, with education as the primary explanatory variable and other demographic characteristics included as controls. The analysis included marginal effects to interpret the impact of each variable, as well as subgroup comparisons by urban status and gender. Further, education was divided into four categories to explore its influence. The results reveal a statistically significant negative association between education and police stops. Specifically, an increase of one year in educational attainment reduces the predicted probability of being stopped by approximately 1.63%. Subgroup analysis confirms that individuals with lower educational attainment, particularly those without a high school diploma, are substantially more likely to report being stopped by police. These results highlight a significant and consistent negative association between education and the likelihood of police stops, suggesting that higher educational attainment may offer protective effects in interactions with law enforcement.

Keywords: education, **police** stops, logistic regression.

1. Introduction

Police stops are a routine but impactful aspect of modern law enforcement. These occur when officers temporarily detain individuals based on suspicion, often without a warrant or arrest. While intended as proactive crime prevention tools, police stops may carry significant psychological and social costs [1]. They might influence individuals' perceptions of justice, erode trust in institutions, and shape long-term attitudes toward the criminal justice system. For many in disadvantaged communities, such stops are frequent and deeply consequential. Given the wide discretion police have in choosing whom to stop, concerns about fairness, bias, and accountability are widespread. Therefore, understanding what factors increase or reduce the likelihood of being stopped is essential—not only for evaluating policing practices, but also for uncovering broader patterns of social inequality embedded in institutional encounters.

Most of the existing research on police stops has focused on visible identity characteristics such as race, ethnicity, and gender. For example, Gelman et al. found persistent racial disparities in stop-and-frisk practices in New York, even after controlling for contextual variables [2]. Epp et al. argued that investigatory stops were often shaped by implicit bias and perceived group identity [3]. Blalock et al. extended this line of inquiry to gender, showing that male and female drivers might be treated differently during traffic stops [4]. While these studies offer valuable insight into the social inequalities embedded in law enforcement, visible traits alone may not account for all disparities. Less visible individual characteristics, such as education, may also influence the likelihood of police contact, but have received relatively little attention. In many studies, education is treated as a control variable rather than a key explanatory factor [1, 2]. Yet education may shape police interactions through multiple ways. For instance, it has the probable ability to increase legal awareness or cause behavioral differences. Lochner and Moretti (2004) found that higher education was associated with lower criminal involvement, suggesting that education may indirectly reduce exposure to policing [5]. However, few studies have directly examined whether education itself reduces the likelihood of being stopped for non-minor offenses. This gap limits the understanding of how structural and individual inequalities intersect in the context of discretionary policing.

This paper aims to investigate whether education influences the likelihood of being stopped by the police for reasons other than minor traffic violations. Using individual-level data from the National Longitudinal Survey of Youth (NLSY) in 1980, this study takes years of schooling as the main explanatory variable and examines its association with police stops. In addition to education, other factors are controlled to reduce potential confounding bias. The main analysis employs a logit model to estimate the probability of being stopped. To further explore possible heterogeneity, this study also conducts subgroup analyses by sex and urban residence. Finally, education is divided into four levels to assess how the relationship may vary across different levels of attainment. By placing education at the center of the analysis, this study offers a new perspective on how non-observable characteristics shape interactions with law enforcement and contributes to broader discussions on the social value of education and institutional equity.

2. Method

2.1. Dataset preparation

This study utilizes data from the National Longitudinal Survey of Youth (NLSY), a nationally representative panel survey that follows individuals over time [6]. The NLSY provides rich individual-level information suitable for examining how non-observable characteristics. This study focuses on the data from 1989, which includes relevant variables related to law enforcement encounters, educational background, and individual characteristics. After processing data, the final sample comprises 4,488 respondents.

The outcome variable of interest is policever, indicating whether an individual has ever been stopped by the police for reasons other than minor traffic violations. The primary explanatory variable is edu (years of completed schooling), which is later divided into four education categories for further analysis. To account for potential confounding factors, several control variables are included: age, gender, marital status, working status, and whether the individual resides in an urban area. Table 1 below presents definitions and descriptive statistics for each variable used in the analysis.

Table 1. Description of Variables Used in the Analysis

Variable	Interpretation
policever	=1 if individual has been stopped by police for non-minor traffic offenses
edu	total years of education (capped at 20 years).
edu_0_11	=1 if has 1ess than 11 years of education
edu_12	=1 if has 12 years of education
edu_13_15	=1 if education between 13–15 years
edu_16plus	=1 if education ≥ 16 years
married	=1 currently married
urban	=1 if in urban location
age	age in years
working	=1 if employed

2.2. Model specification

Given that the outcome is a binary indicator of police stops, a logistic regression model was selected in this study to estimate the likelihood of being stopped by the police. The logit model is appropriate for modeling binary dependent variables and is commonly used in studies examining discrete choices or events [7, 8]. This model estimates the log odds of the outcome as a linear function of the predictors. The general form of the model is given by:

$$logit(P_i) = log\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \cdots + \beta_k \cdot X_k \quad (1)$$

In this model, P_i denotes the probability that individual i has experienced a police stop, while $X_1, \ldots, X_k$ denotes the set of explanatory variables. β_0 is the intercept term, and $\beta_1, \ldots, \beta_k$ are coefficients estimated for each variable.

This model framework enables estimation of how educational attainment and other personal attributes are associated with the likelihood of being stopped, while controlling relevant demographic and socioeconomic factors. The logistic function ensures predicted probabilities remain within the 0–1 range and does not require a linear relationship between covariates and the outcome probability [9, 10].

To facilitate interpretation of the logit model results, marginal effects were computed for key explanatory variables. These effects indicate the change in the predicted probability of being stopped by the police associated with a one-unit change in the independent variable, keeping other variables constant.

3. Results and Discussion

This section presents the results of a series of logistic regression models examining the relationship between educational attainment and the likelihood of having been stopped or recorded by the police (policever). Education (educ) serves as the main explanatory variable, while various demographic and social factors are included as covariates. In addition to the full-sample model, subgroup analyses are conducted by gender, residential area, and education levels.

3.1. Main logit

A binary logistic regression was estimated in the first model, with policever as the dependent variable and educ serving as the primary independent variable. Several covariates, such as age, gender, marital status, and region, were included to control for potential confounding influences.

Table 2. Full-sample Logit Regression and Marginal Effects

Variable	Logit coefficient	Marginal effect	P-value
educ	-0.1193	-0.0163	0.000
sex	1.4661	0.2003	0.000
urban	0.2610	0.0357	0.006
married	-0.3514	-0.0480	0.000
working	-0.1635	-0.0224	0.291
age	0.0046	0.0006	0.799

The results from Table 2 indicate a negative relationship between education and the probability of being stopped by the police. The logit coefficient for educ is −0.1193, suggesting that each additional year of education decreases the log-odds of police contact. The marginal effect of −0.0163 indicates that one more year of education is associated with a 1.63% reduction in the predicted probability of being stopped.

Among the covariates, sex has the largest positive effect, with males significantly more likely to be stopped ($p < 0.001$). Living in urban areas also slightly increases the likelihood of police contact, while being married is associated with a lower probability. However, working status ($p=0.291$) and age ($p=0.799$) do not have statistically significant effects.

This finding suggests that education can be protective against criminal justice involvement, possibly due to its association with stable employment, social capital, and lawful behavior.

3.2. Heterogeneity Analysis

3.2.1. Gender differences

To explore whether the relationship between education and police contact differs by gender, separate logit models were estimated for male and female respondents.

Table 3. Gender-stratified Regression Results

Sex	Logit coefficient	Marginal effect	p-value
Female	-0.1026	-0.0082	0.002
male	-0.1251	-0.0252	0.000

Results from Table 3 shows that among males, the coefficient of educ is -0.1251, with a marginal effect of -0.0252. Among females, the coefficient is -0.1026 with a marginal effect of -0.0082. Both of them are significant.

These results indicate that education is negatively associated with the likelihood of police contact in both genders, but the effect is notably stronger for males. The possible reason is that education may influence behavior and social status more significantly for males, who are more likely to be exposed to policing biases and public scrutiny, leading to a stronger protective effect against police contact.

3.2.2. Region differences

Next, the sample was divided based on whether respondents lived in urban areas to examine environmental effects on the education-police stop relationship.

Table 4. Urban vs. Rural Regression Results

Urban	Logit coefficient	Marginal effect	p-value
urban	-0.1171	-0.0164	0.000
rural	-0.1290	-0.0164	0.000

Given data from Table 4, in urban areas, the coefficient of educ is -0.1171, with a marginal effect of -0.0164, while in rural areas, the coefficient is slightly stronger at -0.1290, with the same marginal effect. Both coefficients are statistically significant at the 1% level.

This suggests that higher education is associated with a reduced likelihood of being stopped by police in both contexts. Although the coefficients are slightly more negative in rural areas, the similarity in marginal effects indicates a relatively consistent pattern across different residential settings. Thus, education appears to display a protective effect regardless of geographical environment.

3.3. Categorical analysis of education levels

To further explore how specific education levels affect the probability of being stopped by police, the continuous education variable (educ) was replaced with four categorical groups: (1) Less than high school (0–11 years) (2) High school graduate (12 years) (3) Some college (13–15 years) (4) Bachelor's degree or above (16+ years, reference group)

Table 5. Education Level Categories and Logit Results

Education	Logit coefficient	Marginal effect	p-value
edu_0_11	0.8647	0.1180	0.000
edu_12	0.7277	0.0993	0.000
edu_13_15	0.3688	0.0503	0.009
edu_16plus (reference group)	0.000	-	-

In terms of marginal effects shown in Table 5, individuals who did not complete high school are 11.8 percentage points more likely to have been stopped by police compared to the college-educated reference group. High school graduates face a 9.9 percent higher risk, while those with some colleges are 5.0 percent more likely to be stopped. These findings emphasize that not only the presence of education but its level matters significantly in reducing police contact, with a clear protective threshold at the level of a completed college degree. The possible reason is that higher education may enhance social awareness, communication skills, and perceived legitimacy, reducing risky interactions with law enforcement, while lower education levels may correlate with socioeconomic disadvantages linked to increased police scrutiny.

These results underscore the significance of achieving a higher education in mitigating the likelihood of police interaction. The disparity between those with partial or lower education and those with a degree highlights education as a central factor in social protection.

4. Conclusion

This study set out to explore how education shape the likelihood of being stopped by police. Using nationally representative data from the 1989 NLSY and a logit model framework, this analysis highlights education as a significant protective factor against police contact. The results show that each additional year of education reduces the predicted probability of being stopped, and individuals without a high school diploma face a substantially higher risk compared to those with a college degree or above. While this study contributes to understanding the structural role of education in discretionary policing, it is not without limitations. The cross-sectional design restricts causal interpretation, and unobserved factors may still influence outcomes. Future research could employ longitudinal methods and consider broader contextual variables to deepen this line of inquiry.

References

[1] K. Petersen, D. Weisburd, S. Fay, E. Eggins, L. Mazerolle, Police stops to reduce crime: A systematic review and meta-analysis. Campbell Systematic Reviews, 19 (1), e1302 (2023).

[2] A. Gelman, J. Fagan, A. Kiss, An analysis of the New York City Police Department's "Stop-and-Frisk" policy in the context of claims of racial bias. Journal of the American Statistical Association, 102 (479), 813–823 (2007).

[3] C. Epp, S. Maynard-Moody, D. Haider-Markel, Pulled Over: How Police Stops Define Race and Citizenship, University of Chicago Press, Chicago (2014).

[4] G. Blalock, J. DeVaro, S. Leventhal, D. H. Simon, Gender bias in power relationships: evidence from police traffic stops. Applied Economics, 43 (29), 4469–4485 (2011).

[5] L. Lochner, E. Moretti, The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. Am. Econ. Rev. 94 (1), 155–189 (2004).

[6] National Longitudinal Surveys (NLS), Access and Data Investigator, https://www.nlsinfo.org/content/access-data-investigator (1989).

[7] D. W. Hosmer Jr, S. Lemeshow, R. X. Sturdivant, Applied Logistic Regression, John Wiley & Sons, New York (2013).

[8] A. Demaris, Logit modeling: Practical applications, Vol. 86 Sage (1992).

[9] M. P. LaValley, Logistic regression. Circulation 117 (18), 2395–2399 (2008).

[10] D. W. Hosmer Jr, S. Lemeshow, R. X. Sturdivant, Applied logistic regression, John Wiley & Sons (2013).