STAT 7760 Spring 2023, Project 1

NBD Count Data of Chicago Police Complaints

Abstract

This paper examines patterns of police misconduct in the Chicago Police Department by analyzing data from citizen complaints filed against CPD officers between the years 2001 and 2010. We fit multiple negative binomial distribution models, using different methods, and focusing on individual propensities. Our analysis provides insight into the the heterogeneous behaviors of police officers and makes inferences that have important implications for improving police accountability and public trust in law enforcement.

1 Introduction

Police misconduct has long been a contentious issue in the United States, with allegations of abuse, excessive use of force, and discriminatory practices being reported regularly. Despite efforts to improve police accountability and transparency, instances of police misconduct continue to occur, highlighting the need for ongoing scrutiny and reform. One way to assess police misconduct is by examining citizen complaints filed against police officers. Complaints provide a window into the behavior of police officers and can reveal patterns of misconduct that might not be captured by official reports or statistics.

The city of Chicago has been a focus of attention for police misconduct in recent years, with high-profile cases of police abuse garnering national attention. In response, the Chicago Police Department (CPD) implemented various reforms to improve its practices, including creating an Early Intervention System (EIS) to identify officers who may be at risk of engaging in misconduct. However, questions remain about the effectiveness of these reforms, and there is a need to continue monitoring the CPD's performance.

This paper aims to answer a variety of managerial questions by analyzing data from citizen complaints filed against CPD officers over a number of years. Specifically, we will investigate the underlying propensities of individual officers to engage in misconduct, drawing conclusions by looking at trends across different years. By analyzing the data in this way, we hope to identify patterns in the behavior of officers and to inform managerial decisions aimed at reducing misconduct and improving public trust in law enforcement. Ultimately, this research has important implications for police accountability and reform, and could contribute to ongoing efforts to improve law enforcement practices in Chicago and beyond.

2 Data

We used data from the Chicago Police Data Project, a living repository of public data about Chicago's police officers and their interactions with the public. The repository contains various datasets covering complaints of misconduct, misconduct investigations, use of force reports, awards, promotions, salary, official rosters, and unit assignments over time.

For this study, we focused on count data of the number of complaints received by each officer. To maintain sparsity, we analyzed data at a yearly level, choosing the years 2001, 2005, and 2010. These years were selected because they are evenly spread out and span a total time period of 10 years. By using these years, we were able to analyze trends in police misconduct over time and identify any changes or patterns that emerged across different periods.

The data was pre-processed and cleaned to remove any duplicate or irrelevant entries, and only officers with at least one complaint were included in the analysis. The final dataset contained information on the number of complaints received by each officer in each year, as well as demographic characteristics such as gender, race, and years of service. The dataset was then used to fit negative binomial distribution models to estimate the underlying propensities of individual officers to engage in misconduct.

3 Methods

To capture meaningful information from the data, we fit four separate models, each capturing different aspects of police complaint data.

3.1 Model 1

We fit a time-varying negative binomial distribution (NBD) model on police complaint count data from the year 2001. We consider one year to be our standard unit of time, and we do not add a spike at zero as we do not expect any officers to be hard core good guys (never receive a complaint). Note, that for this dataset, t lies in the range [0,1]. We fit this model on disaggregate data and select values of the parameters r and α that maximize the maximum likelihood estimate (MLE).

3.2 Model 2

We fit a time-varying NBD model on police complaint count data from the year 2005 using the same approach as in Model 1.

3.3 Model 3

We fit a time-varying NBD model on police complaint count data from the year 2010 using the same approach as in Models 1 and 2.

3.4 Model 4

We fit a time-varying NBD model on police complaint count data from the years 2001, 2005, and 2010. As with Models 1-3, we consider 1 year to be a standard unit of time. However, instead of fitting the model using maximum likelihood estimation, we use the Method of Moments (MOM). We choose this method because we want to be able to fit the model even when the data is not comprehensive. Some police departments may not release detailed data, but rather provide summary statistics, making it necessary to make inferences from limited data. The MOM allows us to fit the model and draw meaningful conclusions even with incomplete data.

We will use the results of these models to analyze individual propensities of officers to engage in misconduct and answer important managerial questions.

4 Results

We report the following results for each of the models. We do not include results from the Chi-square test because our datasets are large and the Chi-square goodness of fit does not perform well on large datasets[1].

4.1 Model 1

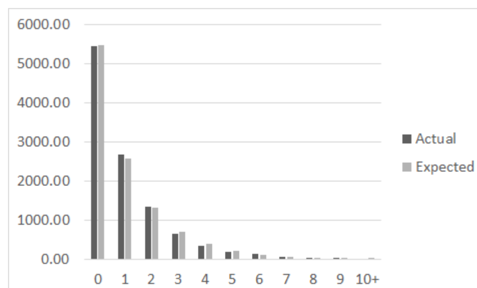
We report the following MLE estimate of parameters r and α for Model 1:

r	α
0.86	0.78

MLE estimates for Model 1

We see that r is small, which indicates that police behavior is heterogeneous. This is to be expected because we know that while many officers are okay, there are some that are really bad. The value of α does not provide any insights.

We now use our obtained parameter to estimate the counts of this dataset. We report the following chart comparing our expected counts to the actual counts:



Expected vs. Actual counts for Model 1

We notice that the expected counts correspond quite well with the actual counts.

Using the insights from the MLE estimates and the comparison between expected and actual counts, we conclude that this model is good enough.

4.2 Model 2

We report the following MLE estimate of parameters r and α for Model 2:

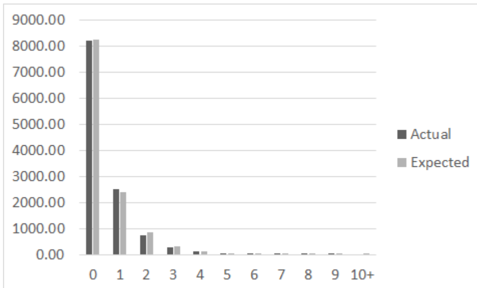
r	α
0.71	1.37

MLE estimates for Model 2

We see that r is small, which indicates that police behavior is heterogeneous. This is to be expected because we know that while many officers are okay, there are some that are really bad. The value of α does not provide any insights.

[1]Quote from STAT 7760 lecture notes: If the dataset is big, we should be skeptical of the chi-square, and use our personal judgments as well as the power of inspecting the histograms to determine the goodness of fit.

We now use our obtained parameter to estimate the counts of this dataset. We report the following chart comparing our expected counts to the actual counts:



Expected vs. Actual counts for Model 2

We notice that the expected counts correspond quite well with the actual counts.

Using the insights from the MLE estimates and the comparison between expected and actual counts, we conclude that this model is good enough.

4.3 Model 3

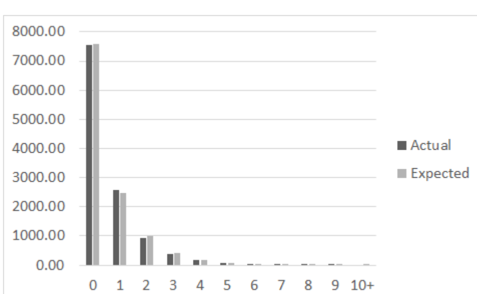
We report the following MLE estimate of parameters r and α for Model 3:

r	α
0.70	1.09

MLE estimates for Model 3

We see that r is small, which indicates that police behavior is heterogeneous. This is to be expected because we know that while many officers are okay, there are some that are really bad. The value of α does not provide any insights.

We now use our obtained parameter to estimate the counts of this dataset. We report the following chart comparing our expected counts to the actual counts:



Expected vs. Actual counts for Model 3

We notice that the expected counts correspond quite well with the actual counts.

Using the insights from the MLE estimates and the comparison between expected and actual counts, we conclude that this model is good enough.

4.4 Model 4

We report the following MOM estimate of parameters r and α for Model 4:

r	α
0.60	0.81

MOM estimates for Model 4

We see that r is small, which indicates that police behavior is heterogeneous. This is to be expected because we know that while many officers are okay, there are a some that are really bad. The value of α does not provide any insights.

Given that we fit this model using the Method of Moments, we cannot compare the expected counts to the actual counts. However, using a sample size of 11,000, we obtain the following expected counts:

x	$\mathrm{E}[x_n]$
0	6805.12
1	2248.47
2	991.05
3	473.28
4	234.72
5	119.00
6	61.21
7	31.80
8	16.65
9	8.77
10+	9.93

Expected counts for Model 4

Using the insights from the MLE estimates, we conclude that this model is good enough.

4.5 Robustness

We would like to highlight the robustness of our models. We cannot make direct comparisons between our models because we fit each on different datasets. However, in each case, we obtain similar estimates, which implies that we are capturing similar propensities across time.

5 Analysis

We will take a deeper look at the models and we will see what conclusions we can make from the individual level propensities.

We will address important managerial questions, such as:

1. Are all police officers bad?
2. Is the police force better than it was 10 years ago?
3. Assuming a police force of 11,000 officers, what do we expect the complaint counts to look like in 1 year? What about in 5 years? 10 years?
4. Across 1 year, what percent of police officers would we expect to have at least one complaint? What about across 5 years? 10 years?
5. Does the 80:20 rule apply to police officers? What proportion of police officers are responsible for less than 10% of complaints? What proportion of police officers are responsible for less than 40% of complaints?

5.1 Models 1-3

Recall that Models 1-3 are time-varying NBD models fit on police count data from the years 2001, 2005, and 2010, respectively. We consider one year to be our standard unit of time. In each of the datasets, t lies in the range [0,1]. We obtained estimates of r and α by maximizing the MLE.

The natural point of discussion is to compare our MLE estimates for each model:

Model	r	α
1	0.86	0.78
2	0.71	1.37
3	0.70	1.09

MLE estimates for Models 1-3

We cannot make inferences based on the value of α, and so we focus our attention on the values of r.

We observe that MLE estimates for r are very similar across Models 1-3, and extremely similar across Models 2-3. This is an interesting result which implies that the individual propensity of police officers to draw a complaint has not changed between the years 2001 and 2010. The low values of r reflect high heterogeneity in officers, which means that officers do not all behave the same.

We do want to provide a possible explanation for the relatively higher estimate of r for Model 1. The complete dataset contains complaints data for all officers who received at least one complaint between the years 2000 and 2016. For the purposes of this paper, we have taken this dataset, and fit each model on officers who were active during that year.

Therefore, Models 2-3 are fit on more officers than Model 1 because we are missing some values for officers who were active during 2001 but did not receive complaints after the year 2000. This results in an under-sample of zero values for Model 1, which leads our model to infer more homogeneity.

5.1.1 Question 1

Are all police officers bad? No!

We notice high levels of heterogeneity in the counts of police complaints, which implies that there is a high spread in the propensities of officers to draw complaints. Some officers have a propensity to draw very few complaints, whereas other have a propensity to draw a lot of complaints. It would be wrong to clump all officer together and claim that they are all bad.

5.1.2 Question 2

Is the police force better than it was 10 years ago? Not really.

There are a variety of ways to define "better": by the average number of complaints, the median number of complaints, the bottom 20% of officers, to name few. In this paper, we define better as being homogeneous. We do not have nearly enough subject knowledge to make an evaluation of how many complaints is considered good versus bad. What we can do, is evaluate whether the police force behaves in a similar manner. If the answer is yes, then we leave it to the experts to determine if on average, officers are good. If the answer is no, then it implies that officers behave differently, and that we have officers with very poor conduct.

Given that all of our estimates for r are low, we can infer that the police force is as, or more, heterogeneous as it was 10 years ago. This leads us to conclude that no, the police force is not better than it was 10 years ago.

5.2 Model 4

Model 4 is a time-varying NBD model fit on police count data from the years 2001, 2005, and 2010.

As with Models 1-3, we consider 1 year to be a standard unit of time. However, instead of fitting the model using MLE, we use the MOM. The main reason for fitting Model 4 with the MOM is that many police departments do not release the full complaints data, and so we want to be able to make inferences even with limited data.

Recall the MOM estimates for r and α:

r	α
0.60	0.81

MOM estimates for Model 4

As with Models 1-3, we note a low value of r which implies a high level of heterogeneity. It is reassuring to observe that even with incomplete data, we can make relatively good inferences.

5.2.1 Question 3

Assuming a police force of 11,000 officers, what do we expect the complaint counts to look like in 1 year? What about in 5 years? 10 years?

We observe the expected complaint counts for 1, 5, and 10 years:

x	1 Year	5 Years	10 Years
0	6805.12	3385.52	2333.77
1	2248.47	1745.36	1293.76
2	991.05	1200.34	956.77
3	473.28	894.40	766.60
4	234.72	692.11	637.90
5	119.00	547.50	542.61
6	61.21	439.38	468.25
7	31.80	356.22	408.22
8	16.65	290.98	358.58
9	8.77	239.09	316.82
10+	9.93	1209.11	2916.73

Expected counts for Model 4

This information is very useful for police departments, governments, organizations, and individuals when trying to use past data to predict future counts. It is remarkable that we are able to use such limited data to obtain these results.

5.2.2 Question 4

Across 1 year, what percent of police officers would we expect to have at least one complaint? What about across 5 years? 10 years?

We observe the implied penetration for 1, 5, and 10 years:

1 Year	5 Years	10 Years
0.38	0.69	0.79

Implied penetration for 1, 5, and 10 years

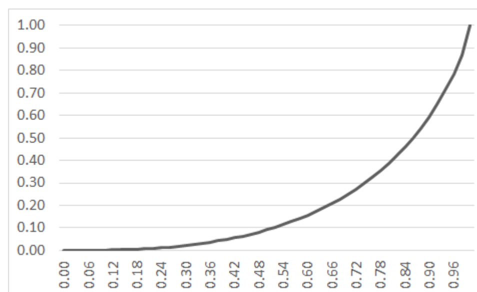
We notice that the implied penetration nearly doubles from year 1 to year 5. However, there is a much smaller increase between year 5 and year 10.

This gives us some sense of hope. We noted earlier that police behavior is heterogeneous. Seeing how the change in implied penetration decreases as the change in time increases gives us hope that there truly are good officers.

5.2.3 Question 5

Does the 80:20 rule apply to police officers? What proportion of police officers are responsible for less than 10% of complaints? What proportion of police officers are responsible for less than 40% of complaints?

We observe the Lorenz curve of police complaints:



Lorenz curve in graphical form

Percentile	L(p)
0.0	0.00
0.1	0.00
0.2	0.01
0.3	0.02
0.4	0.05
0.5	0.09
0.6	0.16
0.7	0.25
0.8	0.39
0.9	0.59
1.0	1.00

Lorenz curve in tabular form

The Lorenz curve of police complaints indicates that police complaints do not follow the 80:20 rule. We notice that officers below the 80^{th} percentile are responsible for 39% of complaints. On the other hand, officers below the 50^{th} percentile are responsible for less than 10% of complaints, and officers above the 90^{th} percentile are responsible for more than 40% of complaints.

These results are both concerning yet also reassuring. It tells us that worst 10% of officers are responsible for more complaints than the best 80% of officers. On the other hand, the best 50% of officers are responsible for less than 10% of complaints.

Take that as you will. As the old saying goes, is the glass half full or half empty?

6 Future Directions

There are several promising avenues for future research on police complaints. Firstly, we suggest making comparisons between the types of complaints filed against officers, the demographic characteristics of complainants and officers, and the outcomes of complaint investigations. This could help us better understand patterns of misconduct and bias within police departments.

Secondly, it would be worthwhile to add additional covariates to our models. For instance, we could include the officer's rank, unit assignment, and prior disciplinary history as predictors of complaint count. This would help us investigate the extent to which these factors influence an officer's propensity to receive complaints.

Thirdly, it is important to consider non-stationarity in police complaint count data. We can investigate whether the patterns we observe in our models persist over time or whether there are changes in the factors that contribute to complaints.

Lastly, we suggest extending this analysis to different police departments. By comparing complaint patterns across departments, we can identify factors that are common or unique to specific contexts. This could help us develop more effective policies and interventions to reduce misconduct within law enforcement.

7 Conclusion

In this paper, we analyzed count data of police complaints to gain insights into the behavior of police officers in a large metropolitan police department. Our analysis revealed that there is substantial heterogeneity in the frequency of complaints across officers, suggesting that some officers are consistently more likely to draw complaints than others. We also found evidence that police officers are less likely to draw complaints as they accumulate more experience on the job.

Our study has several limitations. For example, our analysis is based on data from a single police department, which may not be representative of other departments. In addition, our analysis did not take into account the types of complaints filed against officers or the demographic characteristics of complainants and officers, which could provide important context for understanding our findings.

Despite these limitations, our study provides a starting point for future research on police behavior and public trust. We suggest that future studies could build on our work by considering additional covariates in the models, such as race, gender, and precinct, and examining the outcomes of complaint investigations. Moreover, researchers could compare patterns of complaints across different types of officers and departments to better understand the factors that contribute to police misconduct.

In conclusion, we hope that our study contributes to ongoing efforts to improve policing practices and build trust between law enforcement agencies and the communities they serve. By shedding light on the patterns of complaints against police officers, we hope to inform policies and practices that promote accountability and transparency in law enforcement.