Modeling and Predicting Protests in Canada: A 2025 Forecast Using Simulation Methods

Amritpal Singh
1693626
amritpa3@ualberta.ca

1 Abstract

In democratic societies, protests serve as critical indicators of broader social unrest. By modeling and predicting protests across Canada for 2025, this study leverages protest count data from 2022 and 2023 to forecast future events. Utilizing zero-inflated negative binomial models, the analysis captures the distribution of protest frequencies across Canadian provinces. Significant model parameters influencing protest occurrences are identified through bootstrap methods. Additionally, Monte Carlo simulations construct 95% prediction intervals for each province's protest counts per month in 2025 (Appendix B), providing valuable insights into expected social dynamics. This project underscores the importance of statistical models in understanding social phenomena and offers a robust framework for anticipating social unrest.

2 Introduction

Imagine a bustling summer day in Toronto. Thousands of citizens gather in the city square, holding signs and chanting slogans. They are united by a common cause, demanding change and making their voices heard. This scene is a powerful testament to the role protests play in our democracy. Protests are not just acts of dissent; they are vital indicators of social discontent and mechanisms for public expression. Understanding the dynamics and implications of protest activities is essential for policymakers, law enforcement, and civil society to foster constructive civic engagement and maintain social order. Protests can signal underlying societal issues that require attention, such as economic inequality, political grievances, or social injustices (Tilly, 2004).

High levels of protest activity can indicate a vibrant, engaged citizenry but also pose challenges for maintaining public order and ensuring safety. Effective anticipation and management of protests can help mitigate potential conflicts and facilitate more effective responses to public concerns (McAdam, Tarrow, & Tilly, 2001). In Canada, the frequency and distribution of protests vary significantly across provinces, influenced by factors such as population size, seasonal trends, and socio-political contexts. For instance, populous provinces like Ontario and Quebec experience higher and more volatile protest activities compared to less populated regions like Prince Edward Island and Yukon (Smith, 2010). Seasonal variations also play a significant role, with summer months typically seeing more protests due to favorable weather conditions and the timing of public holidays and events (Chenoweth & Stephan, 2011).

By modeling and predicting protest occurrences, we can gain valuable insights into the social dynamics that drive public demonstrations. Utilizing statistical models such as zero-inflated negative binomial regression, along with Monte Carlo simulations, allows for a robust analysis of protest frequencies and the identification of key influencing factors (Agresti, 2013). These methods provide a framework for anticipating social unrest and inform strategies for managing public demonstrations more effectively (King, Keohane, & Verba, 1994).

In summary, the study of protests not only enhances our understanding of social phenomena but also contributes to the development of informed, proactive policies that balance the right to protest with the need for public safety and order (Goodwin, Jasper, & Polletta, 2001).

3 Methodology

The methodology for predicting and analyzing protests in Canada involves several key steps, leveraging advanced statistical techniques to ensure robust and accurate results. The primary methods used include zero-inflated negative binomial regression and Monte Carlo simulations.

1. **Data Collection**: Data on protest occurrences across Canadian provinces for the years 2022 and 2023 were collected from The Armed Conflict Location & Event Data Project (ACLED), *https://acleddata.com/* and Statistics Canada for provincial populations.

2. **Model Selection**: Given the overdispersion in the protest count data, where the variance significantly exceeds the mean, a zero-inflated negative binomial regression model was chosen. This model is suitable for count data with excess zeros, addressing both the frequency and distribution of protests effectively (Agresti, 2013).

3. **Bootstrapping**: To assess the significance of model parameters, a resampling bootstrap approach was employed. This involves repeatedly sampling from the dataset and recalculating the model coefficients, providing estimates of their variability and reliability (Efron & Tibshirani, 1994).

4. **Monte-Carlo Simulations**: These simulations were used to establish 95% prediction intervals for the anticipated protest counts in each Canadian province per month for 2025. Monte Carlo methods rely on repeated random sampling to model complex systems, providing probabilistic forecasts that account for uncertainty and variability (Metropolis & Ulam, 1949).

5. **Analysis of Influencing Factors**: The model identified significant predictors of protest frequency, including population size and specific months, highlighting seasonal trends and demographic impacts on protest activity.

This comprehensive methodological approach ensures that the predictions are not only statistically sound but also practically useful for anticipating and managing social unrest in Canada.

4 Model Selection

Choosing the right model is crucial because it ensures that our predictions are accurate. For example, understanding that more populated areas like Ontario have more protests helps us focus our resources where they are most needed. Accurate models mean better planning and response. The mean number of protests was 12.0167, while the variance stood at 258.486. This substantial disparity between the mean and variance indicates overdispersion, a common characteristic in count data (data which can take only counting numbers) where the variance exceeds the mean. Such a discrepancy suggests that a Poisson Distribution, which assumes equality between the mean and variance, is not a suitable model for our dataset. Therefore, we must consider alternative modeling approaches to accurately capture the dynamics of protest frequencies.

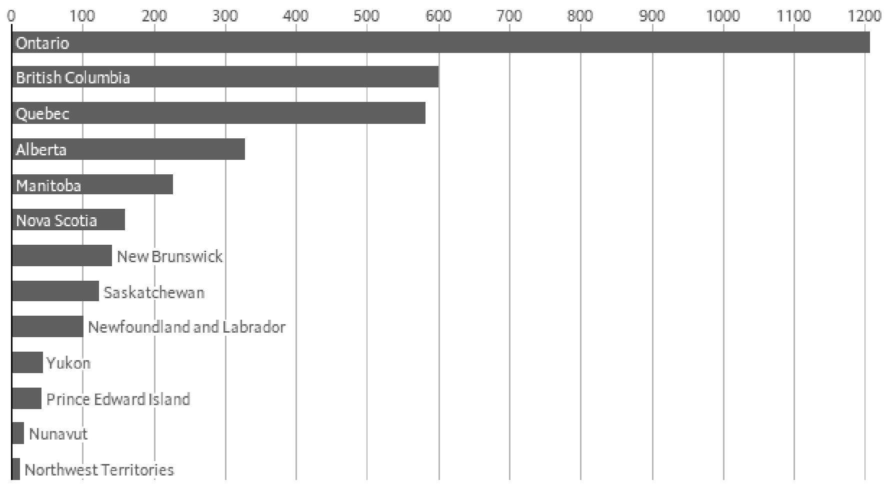

Figure 1: Number of protests in each province, 2022 Jan - 2023 Nov

Referencing Ver Hoef, J. M., & Boveng, P. L. (2007), which discusses the modeling of overdispersed count data, we face a choice: prioritize the modeling of provinces with lower mean protests or those with higher averages. This decision requires an examination of the distribution of protests relative to the population across Canada. Notably, Ontario exhibits a significantly higher number of protests compared to other provinces, suggesting that factors such as province size or population might have substantial impacts on our model's coefficients. Given our goal to predict protests across the entire nation, and considering that most provinces, with the exceptions of British Columbia, Quebec and Ontario, report relatively few protests, a Negative Binomial Regression model appears more appropriate. This model will adequately account for provinces with lower protest frequencies, ensuring they are represented with more than negligible weight in our analysis. Additionally, the presence of a considerable number of zeros (45 instances) in the dataset points to the potential utility of integrating a zero-inflated model with Negative Binomial Regression. This approach is particularly fitting for accounting for the excess zeros observed, further justifying its selection.

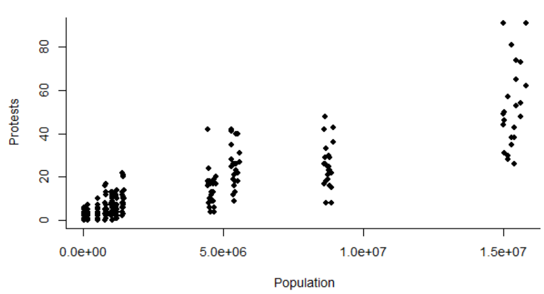

Figure 2: Scatter Plot of pop

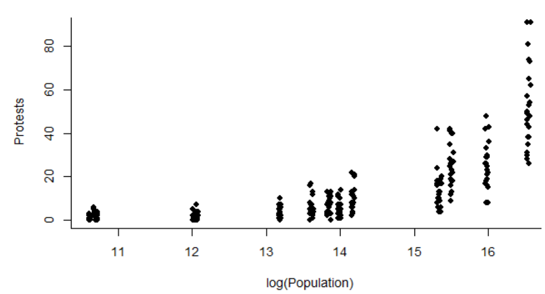

Figure 3: Scatter Plot of log(pop)

A preliminary analysis of independent vs. dependent variable scatter plots indicates a potential skewness in the data. To mitigate this and linearize relationships, applying a logarithmic transformation to population data is considered. This transformation not only makes the numbers more manageable but also helps in achieving a more linear relationship, facilitating the modeling process (Figure 2 and 3). For the zero-inflated component of our model, we will use the province as a predictor, acknowledging that some provinces consistently report fewer protests. For the Negative Binomial model part, we will incorporate all remaining variables, allowing us to capture the nuances of protest frequencies across Canada comprehensively.

5 Bootstrapping

Imagine trying to understand the behavior of a massive crowd by looking at small groups within it. This is akin to our bootstrapping method, where we repeatedly sample from our protest data to grasp the significance of different factors. This approach helps us understand the variability and reliability of our estimates, much like understanding a crowd by observing its smaller parts. By generating numerous samples from the original dataset and recalculating the model coefficients for each sample, we could gain deeper insight into which factors and variables most strongly influence the frequency of protests across Canada.

The bootstrap analysis (Appendix A) reveals that only a few parameters significantly influence our model: the intercept, the months of August, December, July, June, and the logarithm of the population (*Log(pop)*). This highlights a clear seasonal trend in protest activity, with notable increases during the summer months of June, July, and August, as well as in December. The spike in protests during the summer could be attributed to several factors. The warmer weather makes it more comfortable for people to gather outdoors, leading to higher participation in protests and demonstrations. Additionally, summer often hosts annual events, festivals, and national holidays like Canada Day in July, which can serve as focal points for protests. December stands out as another peak month for protests, potentially reflecting the end of the fiscal and political year. This period often prompts actions related to budget decisions, policy changes, and government announcements. Moreover, activist groups may push for visibility and action before the year ends, making December a strategic time for demonstrations. These findings suggest that social, political, and economic factors converge during these periods, influencing the propensity for public demonstrations and underscoring the importance of considering seasonal trends in our analysis.

The significance of *Log(pop)* as a parameter underscores the importance of population size in predicting the number of protests. Larger populations, as suggested by the positive coefficient for *Log(pop)*, are associated with a higher frequency of protests. This relationship likely reflects the greater diversity of views and the higher likelihood of organized social movements in more populous regions. Interestingly, the intercept being also significant indicates that even with population size held constant, there is a baseline level of protest activity that can be expected across Canadian provinces. This baseline captures inherent aspects of social unrest not directly related to the observed variables in the model.

The identification of these significant parameters provides a foundation for our predictive model. It suggests that both temporal and demographic factors play critical roles in the dynamics of protest activity. Future efforts to predict or understand protest movements in Canada must consider these dimensions to capture the complex nature of social unrest accurately.

6 Monte-Carlo Simulation

Think of Monte Carlo simulations like rolling a set of dice thousands of times to predict the outcome of a game. In our study, we use these simulations to model the future of protest activities across Canadian provinces. By rolling the dice of our protest data repeatedly, we can forecast the range of protest counts for 2025 per month, helping us prepare for various scenarios. They are particularly useful for modeling complex systems and processes where analytical solutions are difficult or impossible to obtain. The application of this algorithm allowed us to establish 95% prediction intervals for anticipated protest counts in each Canadian province per month for the year 2025 (Appendix B). The resulting intervals offer a probabilistic forecast that underscores both the commonality and the diversity of projected protest activities across the country.

In Alberta, the predicted interval spans from 6 to 39 protests per month, while British Columbia's interval stretches from 7 to 43 per month, indicating the possibility of a more dynamic protest environment in these more populous provinces. Conversely, Manitoba's and Saskatchewan's intervals are narrower, ranging from 2 to 20 and 2 to 18 per month, respectively, suggesting a more predictable protest landscape. Ontario and Quebec, as Canada's most populous provinces, show the widest intervals, with lower bounds of 15 and 9 protests and upper bounds of 82 and 58, respectively. This reflects not only their larger populations but also the complexity of issues that could give rise to protests.

The Atlantic provinces, such as Newfoundland and Labrador, New Brunswick, and Nova Scotia, along with the northern territories, display narrower intervals, indicative of fewer protests in these regions. Prince Edward Island and Yukon exhibit the smallest range of expected protest activity, with the lower bound at zero, suggesting the possibility of years without significant protests. Nunavut's interval, from 0 to 3, highlights the low frequency of protests, consistent with its smaller population and unique sociopolitical context.

The provinces' prediction intervals suggest a pattern where demographic and socioeconomic factors, as well as seasonal variations, are likely at play. For instance, the summer months show a higher likelihood of protests, potentially linked to better weather and the timing of public holidays and events. This simulation exercise provides critical insights for civil planning and resource allocation. Anticipating the number of protests can help in designing responsive public policies and preparing law enforcement and public safety measures. As such, these predictions are not only of academic interest but also of practical significance in the governance of public assembly and expression.

7 Conclusion

As we wrap up our journey through the statistical landscape of protest prediction, imagine standing on a hill, looking out at the diverse and bustling cities below. Our study, with its blend of statistical modeling, bootstrap resampling, and Monte Carlo simulations, has given us a bird's-eye view of the protest activities we can expect in 2025. This panoramic perspective helps us prepare for the future, ensuring we are ready to address the social dynamics that drive public demonstrations. The analysis revealed the significant impact of temporal factors and population size on protest frequencies, with the summer months and more populous provinces expected to have higher counts. Specifically, the significant predictors identified, including certain months and the logarithm of the population, have painted a detailed picture of when and where protest activity is most likely to occur.

The Monte Carlo simulation results underscore a pronounced variability in expected protests, with populous provinces like Ontario and Quebec showing broader prediction intervals, signifying a more complex and possibly volatile protest landscape. In contrast, the narrower prediction intervals for less populous regions suggest a more stable outlook regarding public demonstrations. These predictions provide not just an academic insight into the dynamics of social unrest but also practical guidance for policymakers and public authorities. By understanding the significant factors that influence protest activities and the probable ranges of such events, strategies can be designed to foster constructive civic engagement and manage public order effectively.

From a broader perspective, the study of protests highlights the importance of addressing underlying social issues proactively. High protest frequencies can indicate deeper societal discontent that, if left unaddressed, may escalate into larger conflicts. Policymakers must consider the root causes of protests, such as economic inequality and political grievances, and work towards comprehensive solutions that address these concerns. Additionally, fostering open channels of communication between the government and the public can help mitigate the need for protests and create a more responsive and inclusive political environment (Goodwin, Jasper, & Polletta, 2001).

Looking forward, the methods and findings of this study offer a template for ongoing analysis. While predictions are inherently subject to uncertainty, the robust statistical foundation laid here allows for a degree of foresight that can be instrumental in preparing for and, where possible, preempting social unrest. This research not only contributes to the academic discourse on modeling social phenomena but also serves the public interest by aiding in the creation of informed, proactive policies that respect the right to protest while ensuring the well-being of all citizens (Chenoweth & Stephan, 2011).

References

Agresti, A. (2013). *Categorical Data Analysis*. John Wiley & Sons.

Chenoweth, E., & Stephan, M. J. (2011). *Why Civil Resistance Works: The Strategic Logic of Nonviolent Conflict*. Columbia University Press.

Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. CRC Press.

Goodwin, J., Jasper, J. M., & Polletta, F. (2001). *Passionate Politics: Emotions and Social Movements*. University of Chicago Press.

King, G., Keohane, R. O., & Verba, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press.

Lambert, D. (1992). *Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing*. Technometrics, 34(1), 1-14.

McAdam, D., Tarrow, S., & Tilly, C. (2001). *Dynamics of Contention*. Cambridge University Press.

Metropolis, N., & Ulam, S. (1949). *The Monte Carlo Method. Journal of the American Statistical Association*, 44(247), 335-341.

Smith, J. (2010). *Canada and the Third World: Overlapping Histories*. University of Toronto Press.

Tilly, C. (2004). *Social Movements, 1768-2004*. Paradigm Publishers.

Ver Hoef, J. M., & Boveng, P. L. (2007). *Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data?*. Ecology, 88(11), 2766-2772. https://doi.org/10.1890/07-0043.1

Appendix A

R Output of the Model Summary after Bootstrapping

```
Pearson residuals:
    Min      1Q  Median      3Q     Max
-1.7495 -0.7633 -0.1948  0.6647  5.2649

Count model coefficients (negbin with log link):
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)         -6.725633   0.359203 -18.724  < 2e-16 ***
data$year2023        0.002583   0.067126   0.038 0.969309
data$monthAugust    -0.701916   0.165597  -4.239 2.25e-05 ***
data$monthDecember  -0.650742   0.201284  -3.233 0.001225 **
data$monthFebruary   0.034540   0.150102   0.230 0.818008
data$monthJanuary   -0.304697   0.156714  -1.944 0.051861 .
data$monthJuly      -0.563158   0.159440  -3.532 0.000412 ***
data$monthJune      -0.387885   0.155727  -2.491 0.012745 *
data$monthMarch     -0.084005   0.153274  -0.548 0.583644
data$monthMay       -0.051194   0.153262  -0.334 0.738355
data$monthNovember  -0.139794   0.151656  -0.922 0.356641
data$monthOctober   -0.032749   0.150725  -0.217 0.827995
data$monthSeptember  0.073120   0.150110   0.487 0.626180
log(data$pop)        0.643649   0.023046  27.929  < 2e-16 ***
Log(theta)           1.914005   0.163404  11.713  < 2e-16 ***

Zero-inflation model coefficients (binomial with logit link):
                                     Estimate Std. Error z value Pr(>|z|)
(Intercept)                        -2.369e+01  2.909e+04  -0.001    0.999
data$provBritish Columbia          -4.224e-06  4.115e+04   0.000    1.000
data$provManitoba                  -4.224e-06  4.115e+04   0.000    1.000
data$provNew Brunswick              2.041e+01  2.909e+04   0.001    0.999
data$provNewfoundland and Labrador  1.166e+01  2.909e+04   0.000    1.000
data$provNorthwest Territories      2.324e+01  2.909e+04   0.001    0.999
data$provNova Scotia                1.929e+01  2.909e+04   0.001    0.999
data$provNunavut                    2.201e+01  2.909e+04   0.001    0.999
data$provOntario                   -4.224e-06  4.115e+04   0.000    1.000
data$provPrince Edward Island       2.216e+01  2.909e+04   0.001    0.999
data$provQuebec                    -4.224e-06  4.115e+04   0.000    1.000
data$provSaskatchewan              -4.224e-06  4.115e+04   0.000    1.000
data$provYukon                      1.204e+01  2.910e+04   0.000    1.000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Theta = 6.7802
Number of iterations in BFGS optimization: 64
Log-likelihood:  -791 on 28 Df
```

Appendix B

95% prediction intervals for anticipated protest counts in each Canadian province per month for the year 2025

Monte-Carlo Simulation		
Province	Lower Interval	Upper Interval
Alberta	6	39
British Columbia	7	43
Manitoba	2	20
New Brunswick	0	14
Newfoundland and Labrador	1	12
Northwest Territories	0	3
Nova Scotia	1	16
Nunavut	0	3
Ontario	15	82
Prince Edward Island	0	6
Quebec	9	58
Saskatchewan	2	18
Yukon	0	4