How Enforcement Start Times Shape Regional Fine Trends: Insights from a Bayesian Hierarchical Analysis*

Earlier Enforcement Start Dates Drive Stronger Increases in Fines Over Time

Qisheng Yu

December 1, 2024

This study examines how the timing of enforcement influences regional fine trends, using a Bayesian hierarchical mixed-effect model to analyze red-light camera data from multiple regions. The findings reveal that regions with earlier enforcement start dates experience greater increases in fines over time, highlighting a significant interaction between enforcement policies and yearly trends. Rigorous model diagnostics and out-of-sample validation confirm the reliability of these results. These insights emphasize the critical role of enforcement timing in shaping compliance behaviors and regional differences, offering valuable guidance for policymakers to design more effective strategies for improving traffic safety and regulatory compliance.

1 Introduction

Running red lights and related traffic violations pose significant risks, contributing to numerous accidents and fatalities annually. Red-light cameras have proven to be an effective enforcement tool, deterring violations and promoting compliance. Studies have extensively examined the general effectiveness of red-light cameras in reducing violations and improving road safety. For instance, Yang et al. (2019) applied propensity score methods to evaluate enforcement impacts on compliance behaviors, demonstrating that red-light cameras reduce violations significantly over time (Yang et al. 2019). Similarly, research by Zhang and Li (2022) examined the interplay between socioeconomic factors and governance policies in shaping urban traffic congestion on a global scale (Zhang and Li 2022). Their study highlights the significant role governance policies play in alleviating congestion, often outweighing the impact of socioeconomic factors. By

---

using a residual trend model, they demonstrated how changes in governance approaches drive variations in congestion patterns over time. While their work provides valuable insights into the macro-level drivers of traffic congestion, it does not address enforcement-specific policies or their direct effects on compliance and fine trends, leaving a critical gap that this study seeks to fill.

Existing studies primarily focus on aggregate measures, such as total reductions in violations or accident rates, often neglecting temporal dynamics and regional heterogeneity in enforcement impacts. For instance, Filtness et al. (2022) conducted a comprehensive review of road safety interventions, including red-light cameras, emphasizing their overall effectiveness in reducing crash rates and improving compliance (Filtness and Williamson 2022). However, the study does not explore how the timing of enforcement implementation or regional disparities influence the long-term trends in compliance and fine patterns. This gap highlights the need for more localized and temporally sensitive analyses, which our study seeks to address. Additionally, most prior work relies on static models, which fail to capture the evolving relationship between enforcement and compliance over time. These limitations highlight a critical gap in understanding the temporal and regional factors that influence enforcement outcomes.

This study seeks to fill this gap by investigating how the timing of enforcement affects regional fine trends and long-term compliance behaviors. Using a Bayesian hierarchical mixed-effect modeling approach, we analyze fines data collected from red-light cameras across multiple regions over several years. This approach allows us to account for ward-level variability through random effects while incorporating interaction terms to examine the interplay between enforcement timing and yearly trends. Unlike previous studies, which often assume uniform effects across regions and time, our analysis emphasizes the dynamic and localized nature of enforcement impacts. In particular, our research analyzes a dataset of fines issued by red-light cameras across various regions over several years (Toronto Open Data 2024). Using a Bayesian hierarchical mixed-effect modeling approach, the study investigates two key questions: (1) How does enforcement start time affect the trend and pattern of fines? (2) To what extent do regional differences shape fine trends? The model incorporates interaction terms to capture the dynamic interplay between enforcement timing, yearly trends, and regional variability, with random effects accounting for ward-level heterogeneity.

The **estimand** of this study is the average change in fines attributable to enforcement timing and annual trends, adjusted for regional variation. This metric provides actionable insights into how enforcement policies shape compliance behaviors over time and across different contexts. The findings reveal that earlier enforcement start times lead to more substantial increases in fines over time, highlighting the importance of timing in enforcement policy design. Furthermore, the results underscore the significant influence of regional factors on compliance patterns, offering a nuanced perspective on enforcement effectiveness that is absent in previous studies.

Our study contributes to the growing literature on traffic safety and enforcement. It provides policymakers with evidence-based insights to optimize the timing and design of enforcement measures, enhancing their impact on road safety and compliance. The structure of this paper

is as follows: Section Section 2 outlines the data and preprocessing steps; Section Section 3 presents the methodology; Section Section 4 details the results; Section Section 5 provides an in-depth discussion of the findings; and the conclusion explores broader implications and potential directions for future research.

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023) for data analysis and visualization. Our dataset (Toronto Open Data 2024) provides comprehensive information on red-light camera fines and enforcement. Following the guidelines of Telling Stories with Data (Alexander and Alexander 2020), we focus on effective communication through meaningful visualizations and statistical insights.

This dataset, sourced from Toronto's Open Data Portal, downloaded through `opendatatoronto` package (Gelfand 2022), provides detailed information on traffic violations recorded by red-light cameras across Toronto from 2007 to 2024. The dataset captures the annual number of fines issued at each camera location, along with corresponding metadata such as the enforcement start date and ward-level information. Red-light cameras are a critical component of Toronto's Vision Zero initiative, aiming to reduce traffic-related injuries and fatalities by improving compliance with traffic laws. The data offers valuable insights into enforcement patterns and trends over time, enabling an analysis of the program's effectiveness and potential disparities across regions.

The dataset consists of multiple variables, each offering unique insights into the dynamics of red-light enforcement across the city:

Variable	Description	Type	Example Values
Location_Code	Unique numeric identifier for each red-light camera location.	Numeric	2501, 2502, 2503
Ward_Number	The municipal ward where the red-light camera is installed (1 to 25).	Character	1, 12, 25
Location	Intersection or address of the red-light camera location.	Character	"Richmond St. and Parliament St."
Enforcement Start_Date	The date when red-light enforcement began at a specific location.	Date	2007-11-09, 2008-03-17
Year_2007 to Year_2024	Annual number of fines issued by the red-light camera for each year.	Numeric	0, 46, 222

Variable	Description	Type	Example Values
Fines	Derived fines for a specific Ward_Number and Year in the long-format dataset.	Numeric	123, 456, 789
Enforcement	Binary variable indicating if enforcement was active (1) or not (0).	Binary (0/1)	0, 1

Table 1 Key Variables in the Red Camera Charges Data

Among the variables above, an additional variable, `Enforcement`, was constructed to indicate whether enforcement was active in a given year, based on the recorded `Enforcement_Start_Date`. To facilitate temporal analysis, the dataset was transformed into a long format, where each observation represents a specific ward-year combination, along with its associated fines and enforcement status. The detailed data cleaning is provided in Appendix A.1.

2.2 Alternative datasets

While this dataset provides detailed and specific insights into red-light camera enforcement in Toronto, alternative datasets were considered but ultimately deemed less suitable for addressing the research questions of this study. Similar datasets, such as provincial-level traffic data or Vision Zero datasets from other jurisdictions, were excluded due to their lack of granularity and enforcement-specific details.

For example, provincial-level traffic datasets (Transportation 2024) often aggregate data across larger geographic areas, making it impossible to isolate the effects of specific enforcement locations or to analyze ward-level variations in fines. Additionally, these datasets typically focus on broader traffic metrics such as collision rates or overall violation counts, rather than detailed year-by-year fine trends linked to specific enforcement start dates. While valuable for understanding macro-level trends, such datasets are not equipped to capture the localized effects of enforcement timing that are central to this analysis. Similarly, Vision Zero datasets from other cities, such as New York City (Data 2024) or Chicago (Chicago 2024), offer insights into traffic safety initiatives but often lack the enforcement-specific details present in the Toronto dataset. For instance, these datasets may include the number of violations or accidents but do not provide enforcement start dates or year-by-year fine distributions, which are critical for modeling the interaction between enforcement timing and yearly trends. Furthermore, variations in policy frameworks and enforcement strategies between jurisdictions would make direct comparisons challenging, limiting the ability to draw actionable insights.

The selected Toronto dataset provides several key advantages. It offers granular data at the ward level, including enforcement start dates, detailed yearly fine trends, and precise location

identifiers. This level of detail allows for the construction of a Bayesian hierarchical mixed-effect model that accounts for both temporal and regional variations, enabling robust analysis of the interaction between enforcement timing and compliance behaviors. Moreover, the dataset's focus on a single jurisdiction ensures consistency in policy context and enforcement strategies, avoiding potential confounding effects from cross-jurisdictional comparisons.

In terms of data quality, the Toronto dataset underwent rigorous cleaning, including the removal of duplicate rows, standardization of date formats, and verification of variable completeness. The final dataset is well-structured, free from inconsistencies, and uniquely suited to addressing the specific research questions of this study. By providing a detailed view of red-light camera enforcement trends and their implications for traffic safety, this dataset supports a robust and focused analysis that would not be achievable with more generalized or aggregated datasets.

2.3 Measurement

The dataset represents a structured collection of traffic violations captured by Toronto's red-light camera system, which operates as an automated enforcement mechanism. In the real world, when a vehicle runs a red light at an intersection, sensors embedded in the road trigger the camera system. This system records detailed information about the event, including an image of the vehicle, the state of the traffic light, the time and date of the violation, and the vehicle's license plate number. These details are then processed by enforcement authorities to confirm the violation. Once verified, a fine is issued to the vehicle owner, reflecting a tangible enforcement response to a specific infraction.

Each individual violation generates a fine, but these fines are aggregated annually for reporting and analysis. In the dataset, these aggregations are represented as yearly fine counts for each camera location, spanning the years 2007 to 2024. Each entry in the dataset encapsulates the total number of fines issued at a specific location in a given year, capturing both the frequency of violations and the effectiveness of enforcement.

To expand the dataset's utility and facilitate analysis, we derived an additional variable: `Enforcement`, a binary indicator reflecting whether enforcement was active during a given year. This variable was created by comparing each camera location's recorded `Enforcement_Start_Date` with the relevant year. If enforcement had already started or was active during that year, the indicator was coded as 1; otherwise, it was coded as 0.

To enable temporal analysis and simplify the modeling process, the dataset was transformed from a wide format, where each year was represented as a separate column, into a long format. In this structure, each row corresponds to a specific combination of `Ward_Number`, `Year`, fines, and enforcement status. This transformation mirrors the progression of real-world enforcement data collection, allowing each entry to directly link a location and time period to the associated compliance data.

2.4 Outcome variables

Statistic	Value
Total_Fines	1036212.0000
Mean_Fines	183.9212
Median_Fines	0.0000
Min_Fines	0.0000
Max_Fines	6615.0000

Table 2 Numerical Summary of Outcome Variables Fines

The dataset records a total of 103,621 fines, reflecting the scale of red-light camera enforcement across Toronto over the years. The average number of fines per ward-year combination is approximately 183.92, though the median is 0, indicating that a significant portion of ward-year observations had no recorded fines, likely due to inactive enforcement or low violation rates. The fines range from 0 to a maximum of 6,615, with the highest fines observed at specific intersections that may experience higher traffic volumes or lower compliance rates.
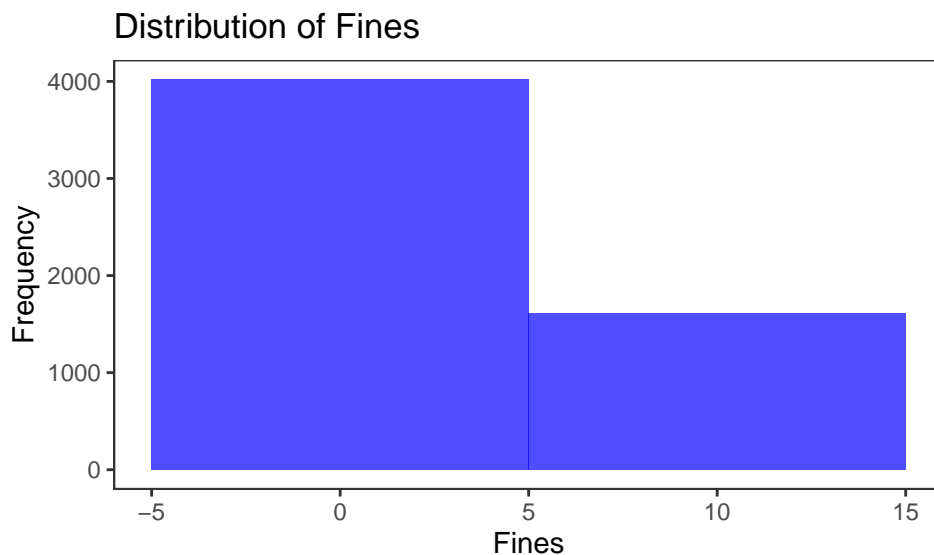

Figure 1: Visualization of Distribution of Fines

Figure 1 shows a highly skewed distribution of fines, with most ward-year combinations clustered around lower values, reflecting inactive enforcement or minimal violations in many cases. The log-transformed x-axis highlights a small number of observations with significantly higher fines, representing intersections with more frequent violations or active enforcement. This distribution emphasizes the need to account for zero-inflation and extreme values in subsequent analyses.
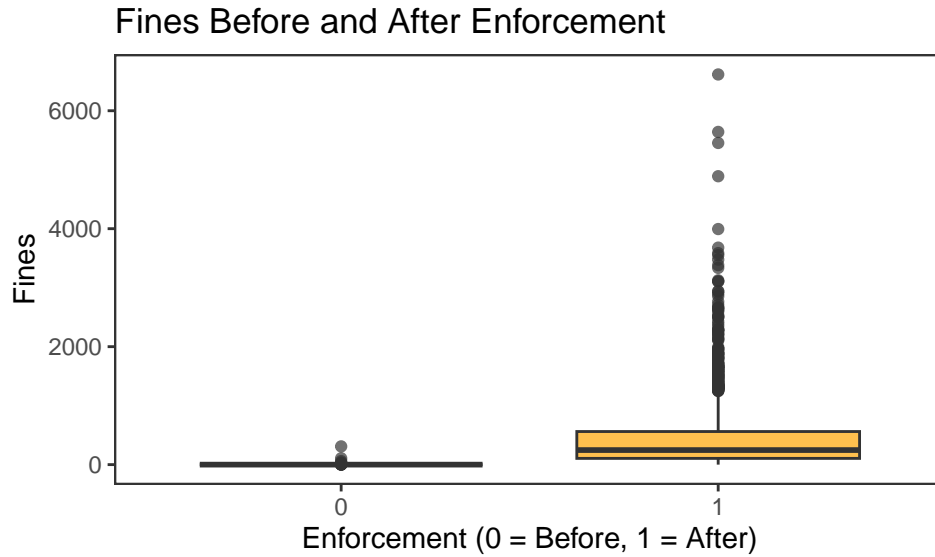
2.5 Predictor variables

2.5.1 Enforcement

Figure 2: Visualization of Distribution of Enforcement

Figure 2 compares fines issued before and after enforcement activation. Observations with `Enforcement = 0` (before activation) show minimal fines, while those with `Enforcement = 1` (after activation) display significantly higher fines and greater variability. This suggests that red-light camera enforcement effectively increases fine issuance, highlighting its impact on traffic violation detection.

2.5.2 Total charges by region

Figure 3 displays the total fines issued across different wards. Wards 10 and 22 exhibit the highest total fines, suggesting concentrated enforcement or higher violation rates in these areas. In contrast, some wards show significantly lower fines, indicating regional variability in enforcement or compliance levels. This highlights potential disparities in traffic patterns or enforcement strategies among wards.

2.5.3 Overall time trend

Figure 4 shows the total annual fines over time, with a clear upward trend peaking in recent years before a slight decline. This suggests an increase in enforcement activity or violations,
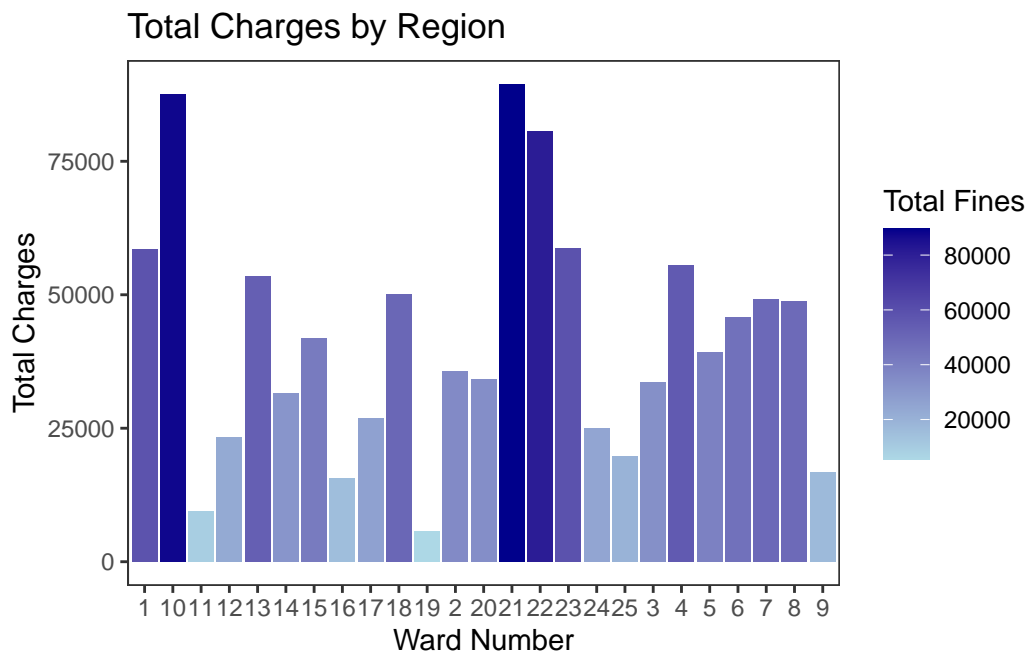
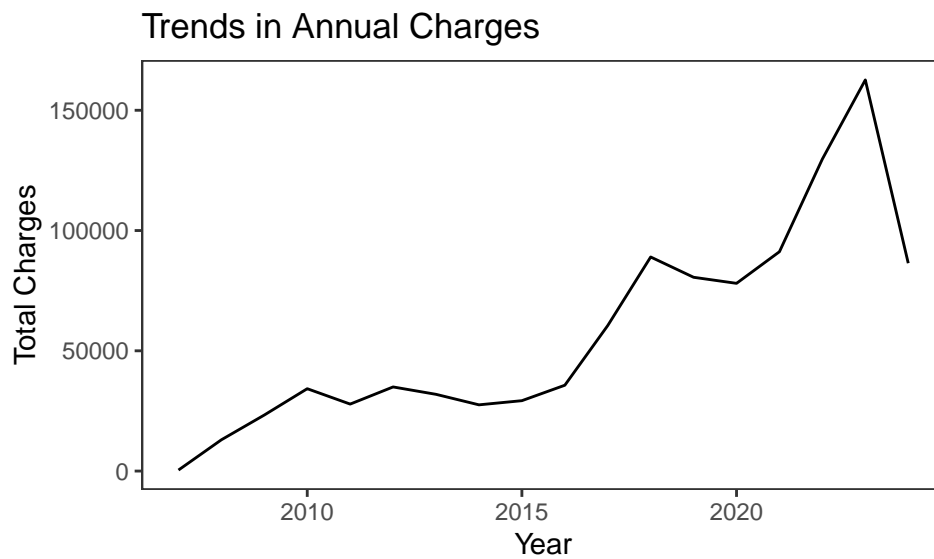Figure 3: Descriptive Statistical Analysis - Total charges by region

Figure 4: Visualization of Time Trends in Charges

followed by a potential reduction, which could be due to improved compliance or other external factors.

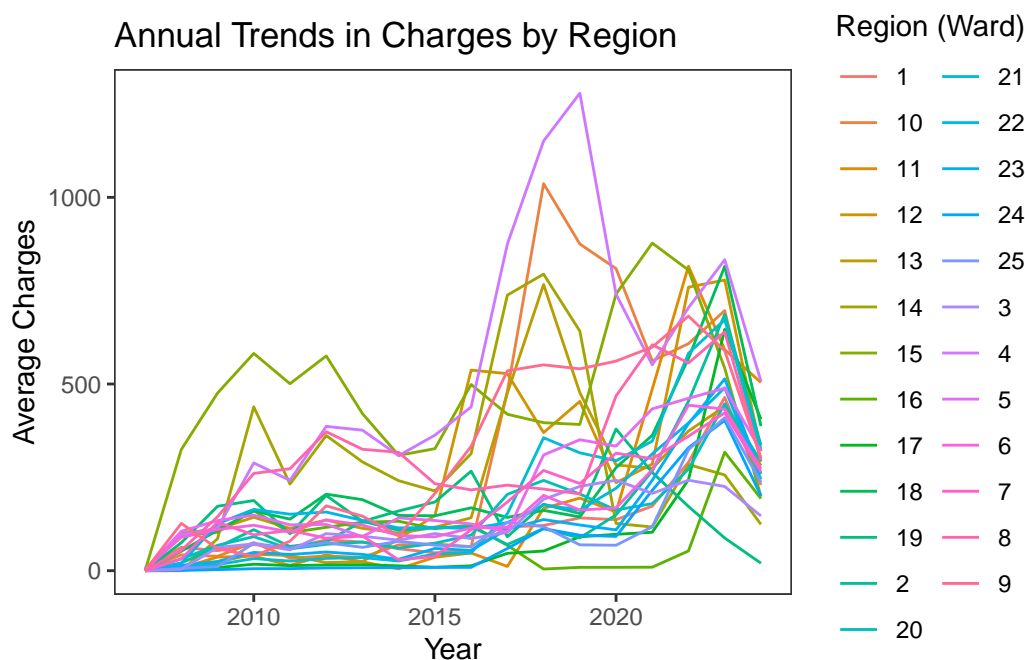### 2.5.4 Time Trends by Region

Figure 5: Visualization of Time Trends in Charges by Region

Figure 5 illustrates annual trends in average charges across regions. While most regions show a gradual increase over time, certain wards, such as 21 and 22, exhibit sharp peaks in recent years. This suggests variations in enforcement intensity or compliance rates between regions.

In the data section, we used opendatatoronto (Gelfand 2022) and dplyr (Wickham et al. 2023) to download data, knitr (Xie 2024) and kableExtra (Zhu 2024) to create tables, tidyverse (Wickham et al. 2019), testthat (Wickham 2011), and arrow (Richardson et al. 2024) for data cleaning, simulation, and testing, and ggplot2 (Wickham 2016) for visualization.

3 Model

The objective of our modeling framework is twofold. First, we aim to explore the differences in trends across regions with respect to red-light camera fines. Second, we investigate the impact of enforcement start time on the observed changes in fines.

To achieve these goals, we employ a Bayesian analysis model, leveraging its flexibility in incorporating prior knowledge and uncertainty. Detailed specifications of the model, priors, diagnostics, and validation techniques are provided in Appendix A.3.

3.1 Model set-up

Define Y_{ij} as the total number of fines issued in year j for region i. We aim to evaluate how fines are influenced by time (Year) and enforcement status (Enforcement), while accounting for variations across regions. The model is specified as:

$$Y_{ij} \sim N(\mu_{ij}, \sigma^2)$$

$$Y_{ij} = \beta_0 + \beta_1 \cdot \text{Year}_{ij} + \beta_2 \cdot \text{Enforcement}_{ij} + \beta_3 \cdot (\text{Year}_{ij} \cdot \text{Enforcement}_{ij}) + u_i + \epsilon_{ij}$$

Where:

- β_0: Overall baseline level of fines, with $\beta_0 \sim N(1000, 500)$.

- β_1: Fixed effect for year trends, with $\beta_1 \sim N(0, 10)$.

- β_2: Fixed effect for enforcement status, with $\beta_2 \sim N(0, 50)$.

- β_3: Interaction effect between year and enforcement status, with $\beta_3 \sim N(0, 10)$.

- $u_i \sim N(0, \sigma_u^2)$: Random effect for ward i, accounting for regional variation.

- $\epsilon_{ij} \sim N(0, \sigma^2)$: Residual error term.

We implement the model using R (R Core Team 2023) and the `brms` (Bürkner 2021) package. The priors reflect weakly informative beliefs, allowing the data to guide parameter estimation while incorporating reasonable expectations based on exploratory analysis.

3.1.1 Model justification

The proposed Bayesian hierarchical mixed-effect model captures both temporal and enforcement effects on the fines. Specifically:

Fixed Effects:

- The inclusion of $\beta_1 \cdot$ Year reflects the overall trend in fines over time, addressing the research question regarding temporal changes in fines.

- The term $\beta_2 \cdot$ Enforcement models the direct impact of enforcement implementation, providing insights into its effectiveness.

- The interaction term $\beta_3 \cdot (\text{Year} \cdot \text{Enforcement})$ allows us to examine whether the enforcement modifies the temporal trend, supporting the investigation of enforcement timing effects.

Random Effects: The random intercepts u_i capture regional variability across wards, allowing us to account for differences not explained by fixed effects.

Priors were chosen to be weakly informative, reflecting reasonable assumptions while letting the data dominate parameter estimation. For example, $\beta_0 \sim N(1000, 500)$ reflects typical fine levels based on exploratory analysis. The hierarchical structure ensures that data from all wards inform the estimation of overall trends while allowing for ward-specific deviations.

This model is well-suited to address the two research questions: (1) identifying temporal trends and (2) understanding the role of enforcement in influencing fines. The Bayesian framework also provides credible intervals, enabling a robust assessment of uncertainty.

3.1.2 Model Decisions: Features

The modeling decisions made in this study closely align with the characteristics of the dataset and the research objectives. Specifically:

- **Inclusion of Year as a Continuous Variable**: The dataset provides annual fine counts from 2007 to 2024. By treating Year as a continuous variable rather than categorizing it, the model captures the overall temporal trends in fines and avoids loss of information that could arise from arbitrary grouping.

- **Incorporation of Enforcement Status**: The `Enforcement` variable, derived from the dataset's `Enforcement_Start_Date`, is included to evaluate the impact of enforcement timing on fines. This binary indicator allows the model to distinguish periods before and after enforcement activation, reflecting a key explanatory factor tied to the dataset's structure.

- **Interaction Terms for Dynamic Analysis**: The interaction term between `Year` and `Enforcement` accounts for the possibility that the effect of enforcement evolves over time. This decision was informed by the observation that fine trends vary substantially across different time periods and enforcement scenarios in the dataset.

- **Inclusion of Ward-Level Random Effects**: The dataset captures data from 25 wards, with notable differences in fine trends across regions. Including ward-level random effects allows the model to account for unobserved heterogeneity across regions, which is critical given the regional disparities highlighted in the dataset.

- **Long Format Transformation and Its Role**: The decision to model fines in a long format, where each row represents a unique combination of `Ward_Number`, `Year`, and corresponding fines and enforcement status, aligns with the need for hierarchical modeling. This structure enables the explicit linkage of temporal, regional, and enforcement effects.

By grounding these modeling choices in the dataset's structure and key features, the approach ensures that the model effectively captures the relationships of interest while addressing the nuances of the data. These decisions also support the research aim of exploring temporal dynamics and regional heterogeneity in fine trends and enforcement impacts.

3.1.3 Assumptions and Limitations

Our model makes several key assumptions that underpin its structure and interpretation. First, it assumes a linear relationship between predictors (Year, Enforcement, and their interaction) and the response variable (Fines). This implies that the effects of predictors remain constant across their range. Additionally, the model assumes independence between observations in different wards, though a random intercept accounts for variability across regions. Furthermore, the model assumes Gaussian-distributed residuals, meaning the errors are expected to follow a normal distribution. Lastly, it assumes homoscedasticity, or constant variance of residuals across all levels of predictors.

Despite its strengths, the model has some limitations. One major limitation is the potential presence of unmeasured confounders, such as socioeconomic changes or other policy variations, that may affect fines but are not included in the model. Additionally, the use of a single random intercept for wards may oversimplify regional variability by assuming all wards share the same variability structure. While the model includes an interaction term (Year:Enforcement), it may fail to account for higher-order interactions or non-linear relationships between variables. Moreover, the model does not consider temporal autocorrelation, which might lead to biased estimates when fines in consecutive years are highly correlated.

The model may not perform well in certain scenarios. For example, if the fines data exhibit high skewness or heavy tails, the Gaussian assumption may not hold, potentially leading to unreliable predictions. Additionally, the model does not account for spatial correlations, such as neighboring wards influencing each other, which might be relevant in studies of enforcement. It may also struggle to capture abrupt changes in fines caused by significant policy interventions that are not included as predictors. Finally, in cases where relationships between variables are non-linear or involve threshold effects, the current model may be inadequate without incorporating more flexible terms, such as splines or non-linear interactions.

3.2 Alternative Models Considered

In addition to the primary model that includes the interaction term $\beta_3 \cdot (\text{Year}_{ij} \cdot \text{Enforcement}_{ij})$, we considered an alternative model without the interaction term. The alternative model assumes that the effects of enforcement and temporal trends are independent, simplifying the structure of the model. This model is easier to interpret and has fewer parameters, reducing the risk of overfitting, particularly in smaller datasets.

However, the alternative model may fail to capture the potential interplay between enforcement timing and temporal trends, which is central to understanding how fines evolve over time in response to policy changes. The inclusion of the interaction term in the primary model allows us to explore whether the temporal effect of fines differs before and after enforcement begins.

The final model was chosen based on the improved explanatory power and theoretical grounding provided by the inclusion of the interaction term. Model comparison metrics, such as the Leave-One-Out Information Criterion, demonstrated better fit for the primary model, justifying its selection for further analysis.

4 Results

Our results are summarized in Table 3.

Table 3

Parameter	Estimate	Est. Error	Lower 95% CI	Upper 95% CI	Rhat	Bulk ESS	Tail ESS
Intercept	792.08	2253.95	-3583.08	5255.50	1	9722	5701
Year	-0.39	1.12	-2.61	1.78	1	9724	5700
Enforcement	0.36	49.79	-94.63	98.52	1	5510	5331
Year:Enforcement	0.21	0.03	0.16	0.26	1	5429	5329
sd(Intercept)	82.31	14.51	58.24	114.87	1	1726	2576
sigma	321.87	3.36	315.41	328.48	1	10891	5758

Table 3 Summary of the Bayesian model

Table 3 provides a summary of the Bayesian model's estimates and diagnostic metrics. The regression coefficients include:

- **Intercept**: The baseline level of fines before considering the effects of year and enforcement. The wide confidence intervals (CI: -3583.08 to 5255.50) indicate considerable uncertainty around the baseline value.

- **Year**: The coefficient for year is -0.39, suggesting a small, non-significant trend in fines over time without enforcement (CI: -2.61 to 1.78).

- **Enforcement**: The effect of enforcement alone is estimated at 0.36 with wide uncertainty (CI: -94.63 to 98.52), highlighting variability in the data.

- **Year:Enforcement**: The interaction term is significant, with an estimate of 0.21 (CI: 0.16 to 0.26). This suggests that the impact of enforcement increases over time, demonstrating a positive adjustment in fines associated with enforcement year-over-year.

The random effect for `Ward_Number`, represented by the standard deviation of the intercept (sd(intercept) = 82.31), captures regional variability. The model residual variance $\sigma = 321.87$ is stable, indicating reasonable model fit.

All convergence metrics and effective sample sizes (Bulk and Tail ESS) suggest the model has converged effectively and produced reliable estimates. These results reveal the importance of the interaction term in capturing enforcement trends, while other coefficients show weaker or non-significant effects.

4.1 Alternative Model

We also consider an alternative model without interaction term. Our results are summarized in Table 4.

Table 4

Parameter	Estimate	Est. Error	Lower 95% CI	Upper 95% CI	Rhat	Bulk ESS	Tail ESS
Intercept	-1802.61	2226.39	-6063.95	2607.64	1	8917	6476
Year	0.90	1.11	-1.28	3.03	1	8905	6450
Enforcement	395.33	11.69	372.36	418.40	1	9670	6552
sd(Intercept)	83.34	14.95	59.38	117.13	1	2064	2680
sigma	322.01	3.34	315.67	328.50	1	14228	6035

Table 4 Summary of the alternative model excluding the interaction term between Year and Enforcement

The alternative model provides an estimation of the fixed effects of `Year` and `Enforcement` on fines without the inclusion of an interaction term. The key results are summarized in the Table 4:

- **Intercept**: The baseline estimate of fines is -1802.61, which is not meaningful in the current context as the confidence interval includes a wide range (-6063.95 to 2607.64), indicating high uncertainty in this parameter.

- **Year**: The estimated effect of `Year` on fines is positive (0.90), suggesting a slight increase in fines per year. However, the wide confidence interval (-1.28 to 3.03) indicates that this estimate is not statistically significant.

- **Enforcement**: The enforcement variable shows a strong positive effect on fines (395.33), with a narrow confidence interval (372.36 to 418.40). This result indicates a significant and substantial increase in fines following the implementation of enforcement.

- **Random Effects**: The standard deviation of the intercept for `Ward_Number` (83.34) suggests some variability in fines across wards. However, the variability is less pronounced than in the full Bayesian model.

- **Sigma**: The residual standard deviation (322.01) indicates the overall variability of fines unexplained by the fixed and random effects.

The alternative model captures the significant positive effect of enforcement on fines but does not account for interactions between enforcement and year. While simpler, this model may overlook nuanced relationships that are better captured by the Bayesian model, which includes the interaction term. The lack of significance for `Year` also suggests that its isolated effect might be limited when not moderated by enforcement.

4.2 Model Comparison

To compare the Bayesian models, we performed Leave-One-Out Cross-Validation, a robust method to evaluate model predictive performance by iteratively leaving out one observation from the dataset and assessing the model's predictive accuracy on the left-out observation. LOO computes metrics such as the Expected Log Predictive Density, which provides a measure of how well the model generalizes to unseen data, and the LOO Information Criterion, which penalizes model complexity while accounting for predictive performance.

Table 5

Model	elpd_loo	SE	p_loo	looic	% Pareto k (Good)	Very Bad Pareto k
Bayesian Model	-32450.7	212.6	46.6	64901.4	100.0%	1
Alternative Model	-32452.3	213.6	46.9	64904.6	100.0%	1

Table 5 Summary of LOO model comparison

For both models, the Pareto k diagnostic was examined to ensure that the models perform adequately for the majority of observations. For both the `bayesian_model` and `alternative_model`, 100% of observations had $k \leq 0.7$ (good performance), with only one observation flagged as very bad ($k > 1$). The elpd_diff metric from LOO comparisons indicates the relative predictive performance of the models. While the differences between the two models were not statistically significant, the interaction term in the `bayesian_model` provided slightly better predictive accuracy than the `alternative_model`.

4.3 Posterior predictive check

The Rhat values, which assess the convergence of the Bayesian model, are all close to 1, with a minimum of 0.9998 and a maximum of 1.0028. These results indicate that the chains have

Statistic	Value
Min.	0.9998
1st Qu.	1.0001
Median	1.0003
Mean	1.0005
3rd Qu.	1.0006
Max.	1.0028

Figure 6: Results of Rhat Summary

converged effectively, as values near 1 suggest good mixing and reliable parameter estimates. The mean Rhat value of 1.0005 reinforces the overall stability and reliability of the model sampling process. Therefore, no convergence issues are detected in the posterior estimates.
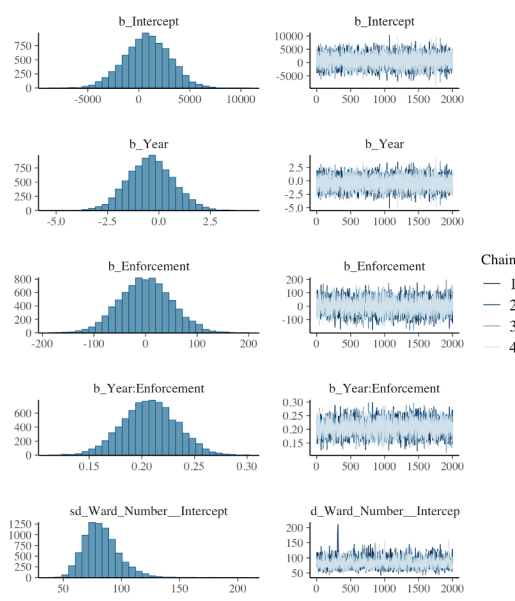

Figure 7: Trace plots and posterior distributions of model parameters

Figure 7 displays both the trace plots and posterior distributions for the key parameters of the Bayesian model. The trace plots (right panels) show the sampling behavior for each chain across iterations, with well-mixed chains and no evident divergences or trends, indicating successful convergence. The posterior distributions (left panels) are smooth and unimodal for parameters such as b_Intercept, b_Year, and b_Year:Enforcement, suggesting stable and reliable estimates. Notably, the posterior for b_Year:Enforcement is tightly concentrated, reflecting high precision in its estimation. The variance observed in parameters like sd_Ward_Number__Intercept aligns with the expected uncertainty in group-level effects.

Overall, the diagnostics confirm the validity of the model's inferences without any significant issues.
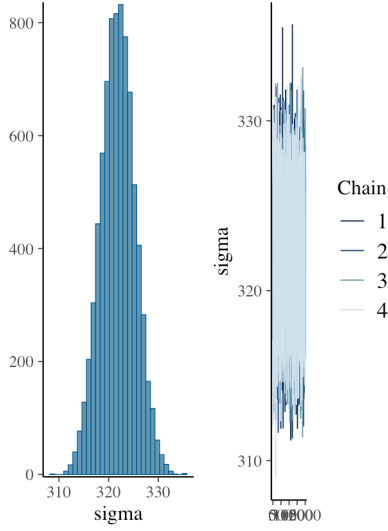

Figure 8: Posterior distribution and trace plot for the residual standard deviation parameter

Figure 8 presents both the posterior distribution (left panel) and the trace plot (right panel) for the standard deviation parameter `sigma` in the Bayesian model. The posterior distribution is tightly concentrated around 321-323, indicating precise and stable estimates of the residual variance in the model. The trace plot displays well-mixed chains with no noticeable patterns or divergence, confirming convergence and effective sampling across all four chains. The density and trace plots together suggest that the parameter `sigma` has been robustly estimated.

Figure 9 shows the posterior predictive check for the Bayesian model, comparing the observed data and the model-generated replicated data. The observed data distribution (dark line) exhibits a sharp peak near zero, indicating a high frequency of small fines, while the replicated data distribution (light line) shows a smoother, more dispersed pattern, failing to fully capture the sharp peak in the observed data. This discrepancy suggests that the model may not adequately account for the extreme concentration of low fine values, indicating room for improvement in model specification.

In the results section, we used brms (Bürkner 2021), caret (Kuhn and Max 2008), Metrics (Hamner and Frasco 2018), and loo (Vehtari, Gelman, and Gabry 2017) for modeling, diagnostics, visualization, and model comparison.
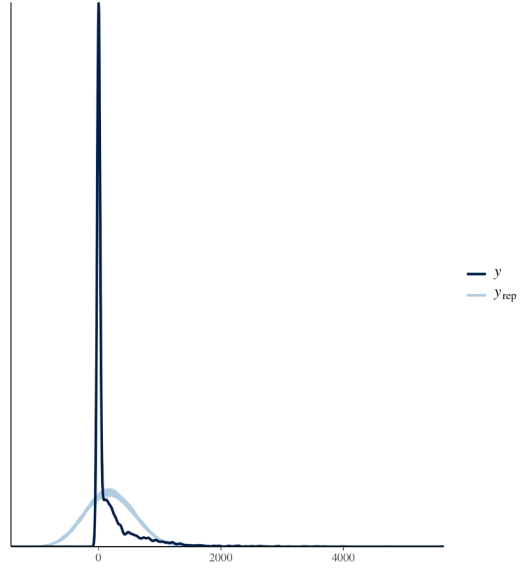
17

Figure 9: Posterior predictive check comparing observed data with replicated data

5 Discussion

5.1 Key Findings and Their Implications

Our analysis demonstrates that the timing of enforcement initiation plays a pivotal role in shaping fine trends across regions. The interaction between year and enforcement provides compelling evidence of how enforcement measures influence the trajectory of annual fines, with this impact gradually intensifying over time. This finding suggests that enforcement policies not only prompt immediate changes in driver behavior but may also exert enduring effects over the long term. Furthermore, the decomposition of random effects by wards highlights the significant influence of contextual factors, such as traffic conditions and socioeconomic characteristics, on fine trends. These results reinforce the importance of considering regional contexts when designing and implementing effective enforcement policies.

The decomposition of random effects by wards further emphasizes the influence of contextual factors, such as traffic conditions, socioeconomic characteristics, or infrastructural differences, in shaping fine trends. This finding is consistent with global research on urban traffic management, which underscores the interplay between governance policies and regional dynamics in influencing compliance patterns (Zhang and Li 2022). These results reinforce the importance of tailoring enforcement strategies to local circumstances, rejecting one-size-fits-all solutions.

5.2 Interpretation of the Bayesian Model Results

The Bayesian hierarchical mixed-effects model employed in this study provided a robust framework for evaluating enforcement policies by distinguishing between fixed and random effects. The posterior analysis revealed that while the overall baseline (intercept) and year trends lacked statistical significance, the interaction between year and enforcement was both statistically significant and positively associated with fines, albeit with a modest effect size. This interaction suggests that enforcement measures may strengthen over time, either by increasing compliance or reducing violations.

Notably, the alternative, simpler model—excluding the interaction term—produced comparable conclusions but demonstrated slightly inferior performance in leave-one-out cross-validation metrics. This underscores the necessity of accounting for interactions to uncover trends that might otherwise remain hidden, as discussed in studies on automated enforcement policies (Yang et al. 2019). This finding highlights the critical role of accounting for interactions in policy impact modeling, as they can reveal nuanced trends that might otherwise be overlooked in simpler frameworks. Furthermore, the Rhat values and effective sample sizes confirmed the convergence of the models, ensuring the reliability and robustness of parameter estimates.

5.3 Limitations and Future Directions

Despite the strengths of our approach, several limitations warrant consideration. First, the sharp peak observed in the fines distribution, which our posterior predictive checks struggled to replicate, suggests potential model misspecification. Future iterations could address this by incorporating additional covariates, such as traffic volume, population density, or socioeconomic indicators, to better account for variance in fines. For instance, literature on traffic congestion highlights the critical role of socioeconomic and governance variables in shaping compliance behaviors (Zhang and Li 2022).

Second, our analysis focuses solely on fines as an outcome measure, which provides only a narrow perspective on enforcement effectiveness. Broader metrics, such as reductions in accidents, improvements in compliance rates, or shifts in public perception, remain unexamined. Including these dimensions in future research would provide a more comprehensive evaluation of the social and behavioral impacts of enforcement policies.

Lastly, while the dataset spans a considerable timeframe, it is geographically confined to a single region. Expanding the analysis to multiple regions or jurisdictions could enhance the generalizability of the findings and offer deeper insights into how regional characteristics and enforcement strategies interact to shape outcomes.

5.4 Policy Implications and Broader Context

Our findings offer valuable guidance for policymakers. The significant year-enforcement interaction highlights the importance of sustained enforcement efforts, as their impact on fines appears to strengthen over time. This underscores the need for a long-term, strategic approach to ensure the enduring success of enforcement policies. Additionally, the substantial random-effect estimates associated with ward-specific intercepts reveal the necessity of tailoring enforcement measures to local contexts. A one-size-fits-all approach is insufficient to address the diverse factors driving violations across regions.

Furthermore, the study underscores the critical role of accurate, granular data in evaluating the effectiveness of enforcement policies. The application of Bayesian hierarchical mixed-effects modeling demonstrates the value of incorporating random effects to capture localized trends. This approach enables policymakers to design more targeted and impactful interventions, offering a nuanced understanding of how enforcement policies can be optimized to address regional variations effectively.

5.5 Methodological Insights and Reflection

This study highlights the significant advantages that Bayesian methods bring to policy research. The flexibility of Bayesian hierarchical mixed-effects models enabled us to account for regional heterogeneity and to investigate complex interactions, such as those between year and enforcement. Moreover, the use of informative priors proved instrumental in stabilizing the model, leading to more reliable parameter estimates, particularly in the context of the considerable variability in fines observed across years and regions. These methodological insights align with best practices in observational data analysis, as outlined in studies on automated enforcement (Yang et al. 2019).

However, the posterior predictive checks revealed some divergence between the model's predicted distributions and the observed data, particularly at the extreme ends of the fine values. This underscores the critical importance of thorough model diagnostics and suggests that alternative modeling approaches, such as those incorporating non-Gaussian distributions, could better capture the nuances of the data. These findings serve as a reminder of the iterative nature of model refinement and the need to explore flexible approaches to improve model fit and predictive accuracy.

5.6 Weaknesses and next steps

In this study, enforcement is represented solely by fines—a tangible yet inherently limited metric. While fines provide a measurable outcome, they fail to capture the broader societal impacts or behavioral shifts resulting from enforcement policies. Future research could address this limitation by incorporating diverse data sources, such as accident reports, compliance

rates, and public opinion surveys, offering a more comprehensive evaluation of enforcement effectiveness and its societal implications.

Another limitation lies in the relative simplicity of our modeling approach. Although the inclusion of interaction terms and random effects added depth to the analysis, more advanced modeling techniques, such as spatial models or time-varying coefficients, could better account for complex dependencies and dynamic relationships. Future studies should embrace these richer approaches to uncover deeper insights into the multifaceted interactions shaping enforcement outcomes.

Lastly, to enhance the generalizability of findings, future research should consider expanding the scope to include multiple regions, varied enforcement policies, and datasets spanning longer time periods. Additionally, exploring how enforcement measures intersect with broader trends, such as advancements in traffic surveillance technology, could provide valuable insights for shaping future policy directions.

A Appendix

A.1 Data cleaning

The raw dataset underwent several cleaning and transformation steps to ensure it was ready for analysis. Column names, originally located in the fourth row, were promoted to the header, and descriptive names were assigned to variables such as `Location_Code`, `Ward_Number`, and year-specific fine counts (`Year_2007` to `Year_2024`). Dates in the `Enforcement_Start_Date` column, initially in Excel's serial format, were converted to standard R date format for consistency, and the irrelevant `Enforcement_End_Date` column was removed.

Yearly fine columns were converted to numeric, with missing values replaced by `0` to account for inactive enforcement or no violations. Data entries containing `*` were cleaned to remove footnotes or annotations, and rows with missing values were dropped to maintain consistency.

The dataset was transformed from wide format into a long format to facilitate temporal analysis. A new binary variable, `Enforcement`, was created to indicate whether enforcement was active (`1`) or not (`0`) in a given year, based on the start date. The cleaned data was saved in both wide and long formats for flexibility in analysis.

A.2 Surveys, Sampling, and Observational Data

A.2.1 Strengths and Limitations of the Dataset

The dataset used in this study stems from observational data collected from red-light camera sites across multiple regions in the United States. Its key strength lies in its near-comprehensive coverage of monitored locations and associated fines issued from 2007 to 2024. This makes it a valuable resource for understanding real-world enforcement practices and their temporal trends. Additionally, the population-level census nature of the dataset reduces sampling bias, as it includes all sites under monitoring during the study period.

However, as is common with observational data, the dataset is subject to several limitations. Notably:

- **Confounding Factors**: External variables such as population changes, economic shifts, or modifications to traffic laws are not explicitly accounted for in the dataset. These variables may distort the relationship between enforcement start times and fine issuance.

- **Selection Bias**: Enforcement initiation was likely not random. High-violation regions may have been prioritized for enforcement, potentially skewing the observed effects of enforcement policies.

- **Measurement Inconsistencies**: Differences in regional definitions of enforcement start dates and data collection practices could introduce variability that complicates direct comparisons.

To illustrate these limitations, a simulated dataset was created, assuming that enforcement tended to be introduced earlier in regions with initially higher fines. The simulation showed that failing to account for this systematic bias could result in overestimating enforcement's effect on fines. This highlights the need for robust statistical techniques, such as propensity score matching or hierarchical modeling, to address confounding and selection biases effectively.

A.2.2 Ideal Data Collection and Sampling

A.2.2.1 Ideal Sampling Strategy

To achieve a more comprehensive and unbiased understanding of enforcement policies, an ideal dataset would rely on stratified random sampling. This approach would involve dividing regions into strata based on relevant characteristics, such as population density, average traffic volume, socioeconomic status, and historical violation rates. A representative sample of sites would then be randomly selected from each stratum to ensure balanced and unbiased data.

Strata Definition:

- **Urban vs. Rural**: Traffic patterns and enforcement needs often differ significantly between urban and rural areas.

- **High vs. Low Violation Regions**: Grouping regions by historical violation rates ensures that both high- and low-violation areas are represented.

- **Socioeconomic Factors**: Strata could include regions grouped by income levels or educational attainment to capture disparities in compliance behavior.

- **Traffic Volume**: Regions with varying average daily traffic counts would be sampled to understand how traffic density influences enforcement effectiveness.

Sampling within Strata:

- Each stratum would be assigned a proportional sample size based on its overall weight in the population.

- Random selection within each stratum ensures representative coverage while minimizing sampling bias.

A.2.2.2 Ideal Data Collection Methods

In an ideal scenario, data collection would involve a combination of automated and manual methods to ensure accuracy and comprehensiveness:

Automated Data Collection:

- Red-light camera systems would be equipped with standardized software to record detailed information about each violation, including timestamps, vehicle speeds, and contextual images.

- Traffic sensors and GPS systems could monitor traffic volume and flow patterns, providing additional context for violations.

Supplementary Data Sources:

- **Survey-Based Metrics**: Public opinion surveys could assess perceptions of enforcement policies, compliance motivations, and general road safety awareness.

- **Crash and Incident Reports**: Integrating data from police reports and insurance claims would provide insights into the broader impact of enforcement policies on road safety.

- **Socioeconomic Data**: Linking traffic data with census data would enable a deeper understanding of how demographic factors influence compliance and violations.

Temporal and Spatial Standardization: Data collection would occur at consistent intervals (e.g., monthly) and across uniformly defined regions to minimize variability and enhance comparability.

A.2.2.3 Ideal Survey Design

A survey could serve as a complementary tool to collect qualitative and quantitative data on enforcement perceptions and impacts. The survey would include:

Target Population:

- Drivers in regions with red-light cameras, stratified by age, gender, and driving experience.

- Local policymakers and enforcement officers to capture their perspectives on policy effectiveness.

Survey Distribution:

- Surveys could be distributed online (via email or traffic-related websites), at DMV offices, or through in-person interviews at traffic hotspots.

Survey Content:

- **Perceptions of Enforcement**: Questions about the fairness, necessity, and impact of red-light cameras.

- **Behavioral Changes**: Self-reported changes in driving behavior since enforcement began.

- **Policy Awareness**: Awareness of enforcement policies and their goals.

- **Compliance Factors**: Factors influencing compliance, such as fear of fines or awareness of road safety benefits.

A.2.2.4 Data Validation and Bias Mitigation

- **Cross-Validation with Independent Data**: Compare red-light camera data with independent sources, such as accident records or insurance claims, to validate findings.

- **Temporal Matching**: Align data collection periods across regions to control for temporal effects.

- **Bias Correction**: Use statistical techniques, such as propensity score matching or instrumental variable analysis, to address confounding and selection biases.

In conclusion, while the current dataset provides valuable insights into enforcement impacts, its limitations highlight the need for more robust and comprehensive data collection methods. An ideal dataset would incorporate stratified random sampling, diverse data sources, and standardized collection practices, enabling a more nuanced understanding of enforcement policies and their broader implications. By addressing these gaps, future research could provide stronger evidence to inform policy decisions and improve traffic safety outcomes.

A.3 Model details

A.3.1 Prior Distribution

Figure 10 illustrates the prior distributions for the beta parameters used in the Bayesian hierarchical mixed effect model, providing a visual representation of the assumed variability and central tendencies before observing the data. Specifically:

- **Beta_0 (Intercept)**: The prior assumes a normal distribution centered at 1000 with a relatively large standard deviation of 500, reflecting a high degree of uncertainty about the baseline fines across wards.
- **Beta_1 (Year effect)**: The prior is centered at 0 with a standard deviation of 10, indicating an expectation of little or no systematic yearly trend in fines, with some allowance for moderate variation.
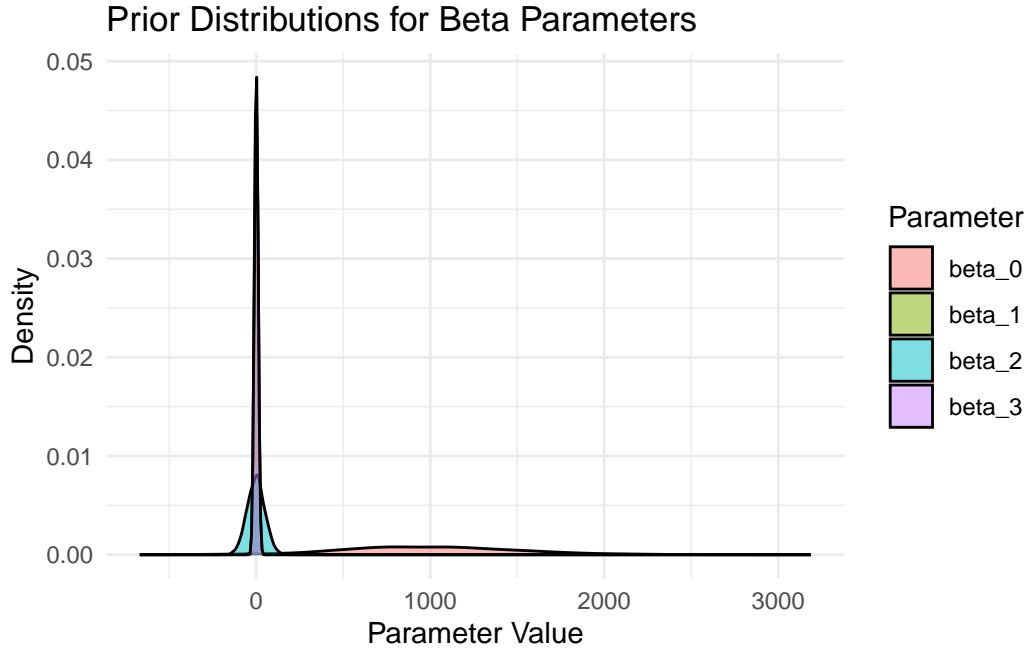
Figure 10: Prior Distribution

- **Beta_2 (Enforcement effect)**: The prior is centered at 0 with a standard deviation of 50, representing moderate uncertainty about the immediate impact of enforcement on fines.
- **Beta_3 (Year × Enforcement interaction)**: The prior is centered at 0 with a standard deviation of 10, suggesting minimal a priori expectation of interaction effects but allowing for moderate variability.

The choice of these priors balances informative guidance (e.g., reflecting plausible ranges based on domain knowledge) and flexibility, ensuring that the data has substantial influence on the posterior distributions. The density plots confirm the expected shapes and spread of these priors, with the intercept exhibiting the widest distribution due to its higher standard deviation.

We used tidyr (Wickham, Vaughan, and Girlich 2023) to convert the data frame to long format.

A.3.2 Diagnostics

Figure 11 demonstrates stable and well-mixed posterior distributions for all model parameters, as evidenced by the smooth and unimodal density plots. The trace plots show good mixing across the four chains with no discernible trends or autocorrelation, indicating convergence of
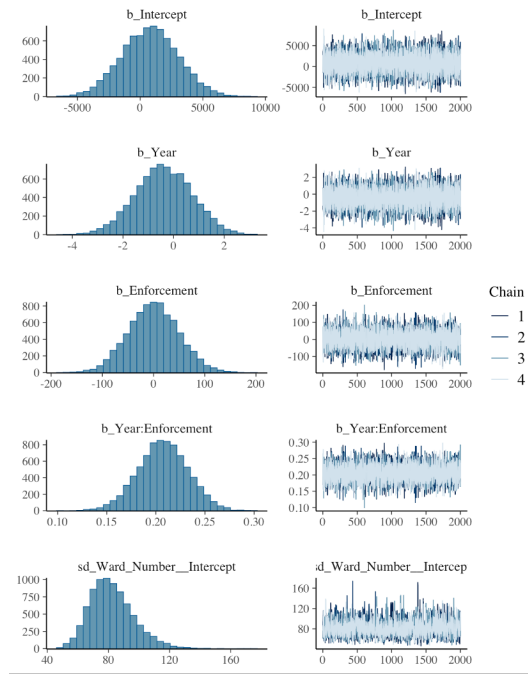
Figure 11: Diagnostic Plot (Posterior Density Plots and Trace Plots)

the Markov Chain Monte Carlo (MCMC) sampling. These diagnostics confirm the reliability of the parameter estimates and support the validity of the model for further interpretation.

Figure 12 for the posterior distribution of σ highlights its stability and convergence. The density plot on the left shows a smooth, unimodal distribution concentrated between approximately 315 and 335, suggesting consistent sampling. The trace plot on the right confirms well-mixed chains without significant autocorrelation or trends, with all chains overlapping and exploring the parameter space effectively. These results indicate that the model's estimation of σ is reliable and robust.

A.3.3 Model Validation and Performance

To evaluate the reliability and predictive capability of the model, we implemented a train-test split where 80% of the data was used for training the model and 20% was reserved for testing. This approach ensures that the model is evaluated on unseen data, reducing the risk of overfitting and providing a realistic assessment of its generalizability. The Root Mean Squared Error, calculated on the test set, was 426.51. This value provides a measure of the average prediction error, with lower values indicating better model performance. While the RMSE is reasonable given the range of fines, it suggests that there is room for improvement, potentially by including additional predictors or using more complex modeling techniques.
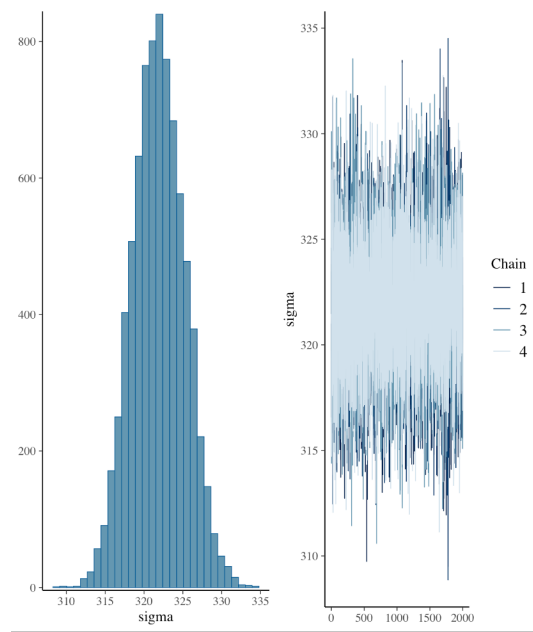
27

Figure 12: Diagnostic Plot (Posterior Density Plots and Trace Plots)

The combination of this validation strategy and diagnostic checks as shown in Section A.3.2 strengthens the credibility of the model's results.

References

Alexander, Rohan, and Emily Alexander. 2020. *Telling Stories with Data*. Self-published. https://tellingstorieswithdata.com/.

Bürkner, Paul-Christian. 2021. "Bayesian Item Response Modeling in R with brms and Stan." *Journal of Statistical Software* 100 (5): 1–54. https://doi.org/10.18637/jss.v100.i05.

Chicago, City of. 2024. "Chicago Vision Zero Crash Data." https://data.cityofchicago.org/.

Data, New York City Open. 2024. "NYC Vision Zero Safety Data." https://opendata.cityofnewyork.us/.

Filtness, Ashleigh J., and Ann Williamson. 2022. "A Systematic Review of Road Safety Interventions: Effectiveness in Reducing Crashes and Violations." *Campbell Systematic Reviews* 18 (3): e1091. https://doi.org/10.1002/cl2.1091.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Hamner, Ben, and Michael Frasco. 2018. *Metrics: Evaluation Metrics for Machine Learning*. https://CRAN.R-project.org/package=Metrics.

Kuhn, and Max. 2008. "Building Predictive Models in r Using the Caret Package." *Journal of Statistical Software* 28 (5): 1–26. https://doi.org/10.18637/jss.v028.i05.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Toronto Open Data, City of. 2024. "Red-Light Camera Annual Charges Dataset." https://open.toronto.ca/dataset/red-light-camera-annual-charges/.

Transportation, Ontario Ministry of. 2024. "Ontario Road Safety Annual Reports (ORSAR)." https://www.ontario.ca/page/road-safety-annual-reports.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2017. "Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC." *Statistics and Computing* 27: 1413–32. https://doi.org/10.1007/s11222-016-9696-4.

Wickham, Hadley. 2011. "Testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

———. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2023. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

Xie, Yihui. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Yang, J. Y., M. Webster-Clark, J. L. Lund, R. S. Sandler, E. S. Dellon, and T. Stürmer. 2019. "Propensity Score Methods to Control for Confounding in Observational Cohort Studies: A Statistical Primer and Application to Endoscopy Research." *Gastrointestinal Endoscopy* 90 (3): 360–69. https://doi.org/10.1016/j.gie.2019.04.236.

Zhang, Y., and X. Li. 2022. "The Relative Roles of Socioeconomic Factors and Governance Policies in Urban Traffic Congestion: A Global Perspective." *Land* 11 (10): 1602. https://doi.org/10.3390/land11101602.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.