UNIVERSITY OF CALIFORNIA
Los Angeles

Testing for Causal Clustering in K-12 Student Discipline

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Applied Statistics and Data Science

by

Sean Mulherin

2024

ABSTRACT OF THE THESIS

Testing for Causal Clustering
in K-12 Student Discipline

by

Sean Mulherin
Master of Science in Applied Statistics and Data Science
University of California, Los Angeles, 2024
Professor Frederic Schoenberg, Chair

The degree to which causal contagion explains the event of student misconduct is investigated. A test introduced by Kresin (2023) and McGovern (2024) is applied, wherein likelihood-ratio tests are performed using information gain statistics to compare the fit of a Neyman-Scott model to that of a Hawkes model. Data records the number of disciplinary actions for K-12 students grouped by day, school, and school year spanning from 2016 to 2023. Evidence of causal clustering appears to vary across grade levels and school years. Out of the twelve different school/year combinations tested, seven exhibited statistically significant evidence of causal triggering. Interestingly, both frequency and magnitude were found to hold strong governance over the conclusion of hypothesis tests.

The thesis of Sean Mulherin is approved.

Michael Tsiang

Nicolas Christou

Robert Gould

Frederic Schoenberg, Committee Chair

University of California, Los Angeles

2024

To God, parents,

and godly parents.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Thank you to all my teachers, coaches, and mentors. Thank you to Rick Schoenberg for providing expert guidance with patience and grace. Thank you to Mike Tsiang for being a kind, supportive, and welcoming teacher; your grace and humility are inspiring. Thank you to Nicolas Christou for your eager support and extensive wisdom. Thank you to Ian McGovern for your sincere generosity and helpful advice. Thank you to Conor Kresin for his work on causal clustering, which forms the basis of the methodology employed in this paper. Lastly, thank you to the Director of Data, Assessment, and Research who generously provided the data.

CHAPTER 1

Introduction

1.1 Background

Education is an integral part of society. Schools provide a space for people to socially, emotionally, and academically develop so that they are better equipped to be productive members of society. Education is so important, most governments mandate at least nine years of compulsory education for all citizens (United Nations Educational, Scientific and Cultural Organization, 2023). According to the Education Data Initiative, the United States provides $810 billion annually to fund K-12 public education, comprising 12.7% of the government's public funding expenditures (Hanson, 2023). The COVID-19 pandemic proved to be pernicious to academic achievement. Prior to the pandemic in the United States, 9 and 13 year-old students were exhibiting slight decreases in reading and math proficiency. This decline was exacerbated by the pandemic, resulting in the largest drop in proficiency scores ever recorded (National Assessment of Educational Progress, 2024). This downward trend can also be seen in older students by observing the average SAT scores in the United States. In 2019, the average scores in math and reading/writing were 528 and 531, respectively. As of 2023, scores dropped to 508 for math and 520 for reading/writing, which equates to 4.13% and 2.11% drops, respectively (Aldric, 2024). Education is not only a crucial aspect of society, but it is also showing alarming signs of hardship. Two main areas of concern have emerged as the most pressing issues in education: student misconduct and student attendance. This paper sets out to provide statistical insight on the former: student misconduct.

To provide insight, a novel point process technique is leveraged to test for causal clustering in the event of student misconduct. This is done both throughout and across school

years. This is a novel research effort in three facets. First, the statistical test used, as proposed by Kresin (2023) and McGovern (2024), has only recently been developed. Second, rigorous statistical research pertaining to modeling student misconduct has historically been neglected. Third, modeling student misconduct by day throughout a school year, as opposed to annually, is an atypical approach; the standard procedure for pedagogical research is to analyze data annually rather than by day. Modeling student misconduct throughout a school year will provide insight into the dynamic nature of the event, equipping educators and administrators alike with information so that they are better able to allocate resources such as professional development, targeted interventions, cultural events, community-building exercises, etc.

The main objective of this research is to test the degree to which causal triggering explains the event of student misconduct. *Causal triggering*, or *causal clustering*, describes a self-exciting phenomenon whereby the occurrence of an event increases the likelihood of, or excites, subsequent events. Note, I will use *disciplinary actions* interchangeably with *student misconduct*. In social sciences, this is often referred to as *social contagion*. Common examples of this social contagion phenomenon are found in yawning and transmissible diseases such as influenza. In the context of student misconduct, if a student is disciplined by a teacher or administrator, causal triggering claims that nearby students are more likely to face disciplinary action as well. There is an important distinction to clarify; students experiencing an increased likelihood of receiving disciplinary action does not necessarily imply that students are more likely to misbehave. Instead, for example, teachers and administrators may be more likely to enforce discipline, somehow becoming more sensitive or confident in their ability to do so.

To test how well causal clustering explains the data, a technique introduced by Kresin (2023) and McGovern (2024) is applied. They propose a likelihood-ratio test to compare the fit of a Neyman-Scott model to that of a Hawkes model. Here, the fitted Neyman-Scott model acts as a baseline to which we can compare a fitted Hawkes model. This works because a Neyman-Scott model disregards causal clustering, whereas a Hawkes model is sensitive to such behavior. If student misconduct is indeed an event induced by causal triggering, a

Hawkes point process model will outperform its non-causal counterpart in explaining the data. Alternatively, if student misconduct is a non-causal event, then these models will offer similar fits. Applying a Hawkes point process to model student disciplinary actions is a notable and novel strategy. Doing so will provide a more in-depth understanding of the nature of social contagion within the context of adolescent misbehavior.

There are two intended audiences for this research. First, I provide educators with concise and straightforward analysis of student misbehavior in recent years. Second, I provide statisticians with rigorous analysis by implementing novel point process techniques. This paper is structured to resonate with both audience groups. I begin in Chapter 2 by explaining the data, continuing on to exploratory data analysis performed with the intention of unearthing notable patterns and trends that will be most helpful for educators. Then, in Chapter 3, I provide brief descriptions of Temporal Point Processes, Hawkes and Neyman-Scott models, and how these models are used in a hypothesis test proposed by Kresin (2023) and McGovern (2024) to test for causal triggering. In Chapter 4, I provide the results of such tests performed on various schools and grade levels, followed by a detailed discussion in Chapter 5. A brief summary, concluding remarks, and future areas of potential research are given in Chapter 6.

1.2 Related Work

Research is underdeveloped in current literature pertaining to tests for causal triggering, with applications of such tests performed sparingly. One study of similar nature to the present research was conducted on criminal activity; Mohler et al. (2011) modeled criminal activity in the region of the San Fernando Valley in Los Angeles, CA using an unmarked spatiotemporal self-exciting point process (see Sections 3.1 and 3.2 for a detailed description of this model). In their study, they used a non-parametric approach that leverages a density estimation procedure to distinguish occurrences caused by exogenous effects from those caused by self-excitation effects. The process of decoupling events in a point process into two categories, exogenous and self-exciting, is referred to as *stochastic declustering* (Zhuang et al., 2002).

Similar work to this can also be found in modeling earthquakes and their aftershocks (Fox et al., 2016), as well as modeling the spread of infectious diseases (Park et al., 2020). This type of statistical inference analysis is referred to as *causal discovery*, or *Granger causality*, and could offer many opportunities to learn more about the nature of social dynamics within a region (Granger, 1969).

Pedagogical research studying K-12 student misconduct is well developed in current literature. A recent study was conducted on disciplinary actions looking at more than 28,000 Pre-K and K-12 public schools in the U.S. (Fabes et al., 2021). This study focused on the practice of exclusionary discipline, defined as punishments that involve the physical removal of a student (i.e. suspension or expulsion). Fabes et al. (2021) noted the enduring harmful effects of excluding students from the school environment after misbehaving, regarding the disciplinary practice as a "significant concern" because it "denies students access to learning and teaching, access to peers, and access to resources that schools provide (Fabes et al., 2021, p. 299-316)." They also noted the disconcertingly high number of exclusionary discipline cases found nationwide as of 2018. This should be taken into consideration when discussing implications of the present research.

CHAPTER 2

Data

The primary data of interest records the daily number of disciplinary actions taken upon K-12 students in a school district. Data is generously provided by the Director of Data, Assessment, and Research from a school district that will remain anonymous. Disciplinary actions include any flagged event that is prohibited according to the school's respective policy. This may include suspension, expulsion, phone misuse, foul language, cheating, etc. This data was collected from an online system used by teachers and administrators to record student misconduct. This tracking is done to provide helpful information about individual students so that teachers, parents, and administrators are better able to make informed rational decisions.

The data span from 2016 to 2023 and are categorized by day, school, and school year. Each school year consists of approximately 175 scheduled school days; although, the COVID-19 pandemic caused discrepancies for the 2019/2020 and 2020/2021 school years in this regard as there are only 137 recorded in-person schools days for these years. After fastidious cleaning, each school year includes data for 7 schools: 4 of which are elementary schools, 1 is a middle school, and 2 are high schools. The school district is in a rural area with a population of approximately 10,000 people, 3,000 of which are students. As such, the extrapolation power of this research is limited.

2.1 Assumptions and Limitations

A few assumptions are made in the forthcoming analysis. As aforementioned, we assume a constant enrollment rate throughout each individual school year but not across school years.

Although it is typical for slight undulations in total class enrollment throughout a school year, it is a marginal amount and so this is a safe assumption to make.

We assume students within the same school are able to witness their peer(s) receiving disciplinary action and thus be exposed to the social contagion imposed by the event. Contextually, witnessing a disciplinary action could come in the form of visual, verbal, or auditory information, eventually resulting in a student-peer's awareness of the event at any time after its occurrence. From this, we will assume that the social contagion of a disciplinary action is contained exclusively within the respective school in which the event occurs.

We assume school policies, particularly those pertaining to disciplinary practices, are consistent across schools and school years. This is a perilous assumption to make because school policies are often susceptible to annual changes and can vary across grade levels. For example, behavior management policies in an elementary school are likely to be different than that of a high school in order to address the needs of a more mature student body. Moreover, school policies are governed by superintendents, principals, vice principals, and teachers, each bringing personal biases, and each having an impactful influence on the disciplinary ecology of a school. As staff positions change, so too do school policies. Indeed, this is a perilous assumption because school policies – specifically how students are punished for their misbehavior – have grave impacts on students' future behavior (Fabes et al., 2021).

Another impactful limitation of this study is unfurnished information regarding the type of misbehavior. We are treating all prohibited behaviors equally. Meaning, for example, cell phone misuse will be treated the same as temporary suspension, which will be treated the same as permanent expulsion, and so on. Misbehavior type was intentionally omitted to ensure data anonymity and protect student privacy. This is different than other studies mentioned in Section 1.2 such as the application of self-exciting point processes on modeling criminal activity, where the type of criminal activity was confined to burglaries and gang violence, independently. Other information such as school policy, school climate, and type of school (i.e charter, private, public) is omitted for this study and thus limits the level of detail we are able to achieve.

2.2 Exploratory Data Analysis

This section includes plots and brief initial discussions on student misconduct data. All the figures in this section were created using all 7 schools within the school district. The initial two figures (Figure 2.1 and Figure 2.2) include all available data from 2016 to 2023. All other figures in this section omit data occurring between 3/1/2020 and 7/1/2021 to account for discrepancies induced by the COVID-19 pandemic. This analysis, just like the data, focuses only on the events within one particular school district. The figures in this section are presented to give the audience an initial overview of the patterns and trends of student misbehavior. Moreover, they are meant to evince the inner workings of the dynamic nature of student misbehavior throughout a school year. This section acts as an exploratory data analysis to set the stage for more rigorous statistical analysis to be performed later on.

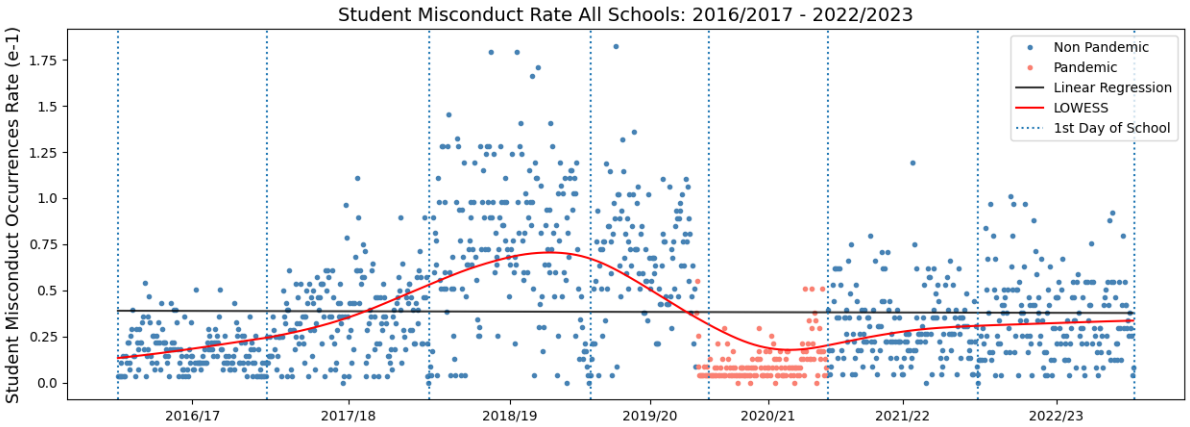


Figure 2.1: Time plot of misconduct occurrences

Figure 2.1 shows a comprehensive view of the data set, portraying the daily student misconduct occurrence rate for all 7 schools in the district for all school days beginning with the 2016/2017 school year and ending with the 2022/2023 school year. Each daily occurrence rate is calculated by taking the ratio of total daily disciplinary case count with total enrollment for the respective school year. A linear regression model (black) is fit to unveil the overall trend with a near zero slope and negative polarity. A smooth trend line (LOWESS) is fit to portray the dynamic behavior of disciplinary occurrences across school years. Notice the steeply increasing trend before the pandemic and relatively constant trend after the pandemic.

I convert the number of daily disciplinary occurrences to the rate of disciplinary occurrences, defined as the ratio between daily occurrences and total annual enrollment for all 7 schools. Annual enrollment data is collected and provided by the state's Department of Education. Daily attendance would provide a more robust rate metric, but such data is not readily available. Therefore, 2.1 assumes a constant enrollment rate throughout each school year.

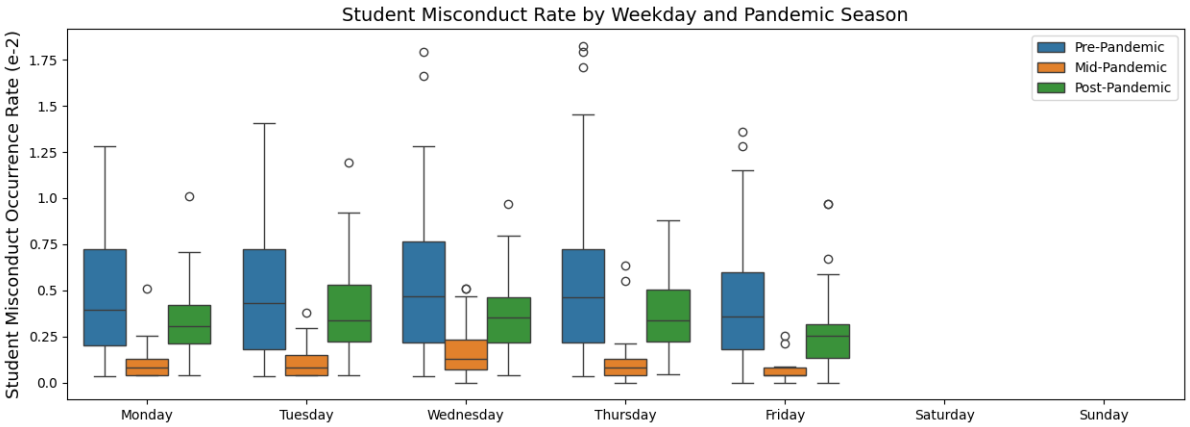


Figure 2.2: Histogram by pandemic

Figure 2.2 portrays descriptive statistics of student misconduct rates by weekday and pre/mid/post pandemic. Pre-pandemic is defined as dates prior to 3/1/2020. Post-pandemic is defined as dates after 7/1/2021. Mid-pandemic refers to all dates between the aforementioned range. Notice the fluctuations in disciplinary rates throughout these periods. Student misconduct was significantly higher before the pandemic compared to after. This decrease in student misconduct after the pandemic could be explained by lower attendance rates after the pandemic. Note, for all succeeding figures in this section, data from the mid-pandemic time span (orange) is omitted.

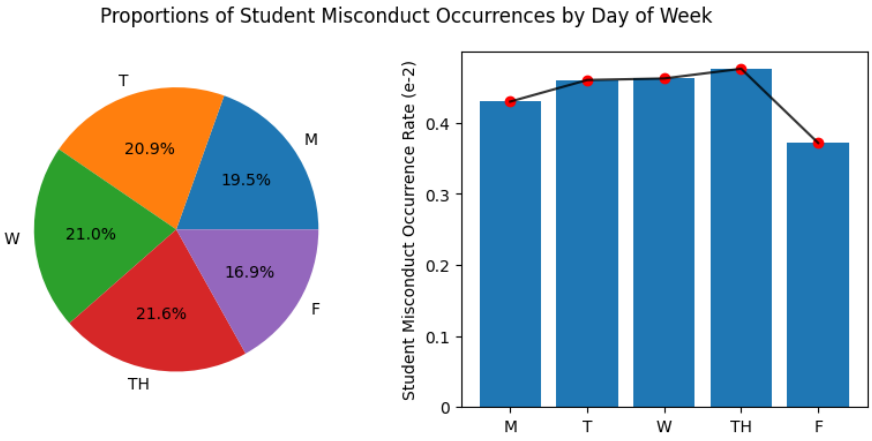


Figure 2.3: Day of week chart

Figure 2.3 portrays the dynamic behavior of student misconduct rates throughout a school week. The mean weekday occurrence rate presented in the bar plot (right) is calculated by grouping by day of week and computing the mean occurrence rate for each weekday. The pie chart (left) takes the ratio of mean weekday occurrence rate with total weekly occurrence rate to obtain each day's proportion of occurrences. For example, given that a student is receiving disciplinary action, there is a 16.9% chance the act occurs on a Friday and a 20.9% chance it occurs on a Tuesday. Notice, students are least likely to receive disciplinary action on Mondays and Fridays. It is possible for this to be explained by a lower mean attendance rate on Mondays and Fridays. Daily attendance rate would provide insight in this regard.

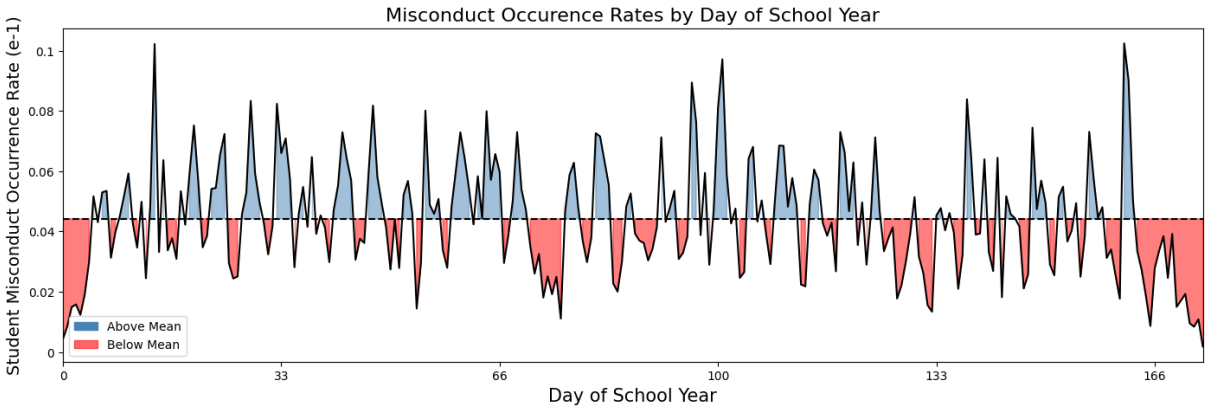


Figure 2.4: Day of year chart

Figure 2.4 portrays the dynamic behavior of student misconduct rates throughout a school year by day and is calculated using the same fundamental techniques as previously described. There are two main breaks throughout a school year: winter break and spring break. Days were removed to account for these breaks in an effort to avoid misleading depictions. Take notice of the prominent maxima and minima that occur throughout a school year. Consider Figure 2.5 below for a broader portrayal.

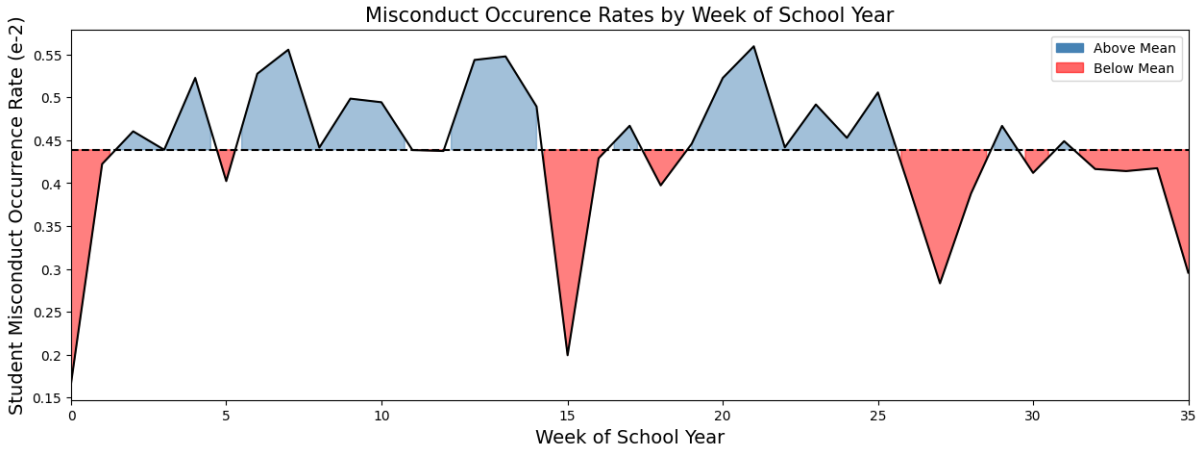


Figure 2.5: Week of year chart

Figure 2.5 portrays the dynamic behavior of student misconduct rates throughout a school year by week. You will notice similar trends as shown in Figure 2.4 but on a weekly scale as opposed to daily. Note, winter and spring break typically happen around weeks 14 and 27, respectively. Here, you can more clearly see four distinct time periods during which students experience the lowest disciplinary action (red). As shown, students receive very little disciplinary action at the outset of the school year, around winter break, around spring break, and towards the end of the school year.

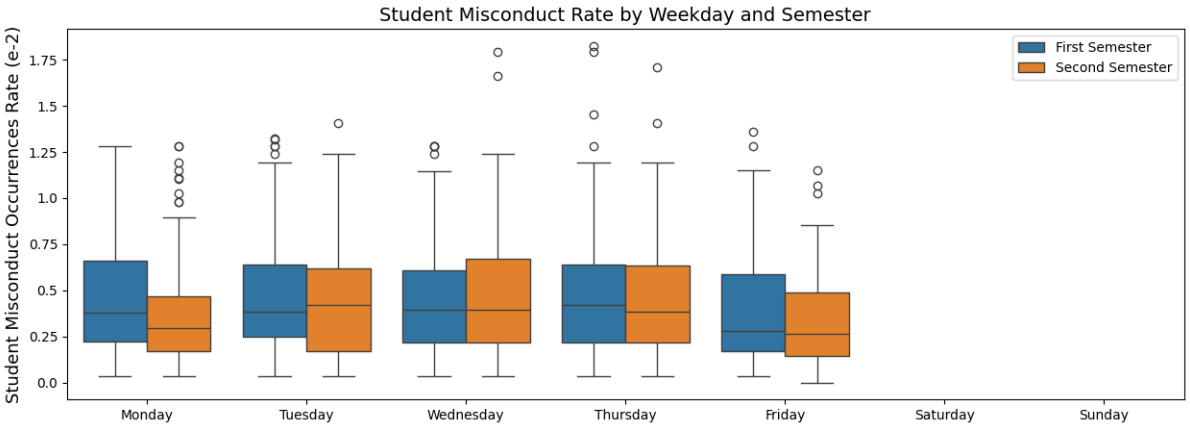


Figure 2.6: Histogram by semester

Figure 2.6 portrays both the central tendencies and measures of dispersion of student misconduct rates by weekday and semester. Notice, students are slightly more likely to experience disciplinary action in the first semester compared to that of the second. A word of caution, as this could potentially be explained by attendance rates if rates typically decrease throughout a school year. If there is a lower overall attendance rate, there are less students available to receive disciplinary action; thus, resulting in a lower rate of disciplinary occurrence.

Overall, as seen in these initial figures, student misconduct rates have decreased from 2016 to 2023. This is a surprising result and a possible explanation, as alluded to earlier, is an increase in student absenteeism. An increase in student absences could explain a decrease in student misconduct because if students are not present, they, of course, are not able to misbehave and receive disciplinary action. Recall, we assume constant enrollment within each school year because daily attendance data is not available. Nationwide, chronic absenteeism, defined as the proportion of students missing 10% or more school days within a school year, has increased at alarming rates in recent years as a repercussion to the pandemic. A recent executive summary by the American Enterprise Institute found that the national average for chronic absenteeism increased from 15% in 2019 to 28% in 2022 and "remained substantially elevated in 2023 (Malkus, 2024)." Future research should be done to analyze the relationship between student misconduct rates and student absenteeism.

This concludes the initial data exploration. A documented code book, as well as a copy of the data with a data management plan can be found at https://seanmulherin.github.io/thesis.html

CHAPTER 3

Methods

A recently proposed test for causal triggering, formulated by Kresin (2023) and McGovern (2024), is employed for the predominant analysis of this paper. The procedure is as follows:

1. Fit a Neyman-Scott model to the data using a Gaussian triggering density.

2. Apply a Gaussian clustering algorithm using maximum likelihood estimation to establish fitted parameters. Namely, μ: number of clusters, A: the mean number of points per cluster, and σ: the standard deviation of each cluster.

3. Simulate 500 Neyman-Scott processes using parameters found in Step 2. For each simulation, fit a Hawkes model using the Expectation-Maximization algorithm and calculate the information gain statistic.

4. Collect all 500 information gain statistics to form a distribution.

5. Calculate the information gain statistic for the original data. Compare it to the sampling distribution to determine if the null hypothesis can be rejected given the predetermined level of significance of $\alpha = 0.05$.

3.1 Temporal Point Processes

In this section, I provide a brief description of temporal point processes. For a more thorough description, refer to Daley and Vere-Jones (2003). A temporal point process is a collection of non-negative integers ($\mathbb{Z}^+$) that describe the magnitude to which an event occurs. In this way, it is known as a counting process wherein $N(t)$ represents the number of points that have occurred at or prior to time t. $N(t)$ is said to be either finitely or infinitely countable and contained within set S, a subset of the non-negative real number line, $\mathbb{R}^+$. As a counting process suggests, $N(t)$ must be a monotonically increasing function that is right-continuous. $f(x)$ is said to be monotonically increasing iff $x \leq y \implies N(x) \leq N(y)\ \forall x, y \in D \subseteq \mathbb{R}$ (Solow, 2013). $f(x)$ is said to be right-continuous if for any x_0 $\lim_{x \to x_0^+} f(x) = f(x_0)$ (Widder, 1989).

Point processes, in essence, model stochastic intervals between successive events and, as such, are particularly suitable for modeling wait times, interarrival times, and other phenomena often regarded by queuing theory, among other research areas (Franken, 1982). As stated, a temporal point process is a two dimensional representation of a countable set of whole numbers that fall on the non-negative real number line, $\mathbb{R}^+$ (Schoenberg, 2011). These numbers are the discrete timestamps of an event's occurrence, acting as a time series. From this, temporal point processes are classified within the broader domain of time-series analysis. A time series describes epochs mapped to $\mathbb{R}^+$, respecting the physical idea that time is continuous.

It should be noted that point processes can be, and often are, generalized to higher dimensions for various purposes, such as *spatial point processes* wherein the location of points in a defined region is modeled, or *marked point processes* wherein auxiliary information is regarded by the model; however, higher dimensional point processes are not the primary focus of the present research. Instead, a strictly temporal point process is used in the forthcoming analysis. This is because the event of student misconduct is assumed to be contained within discrete locations in space (i.e. the respective school within which the event occurs) and shall resemble a strictly temporal point process. Consider a formal definition of a temporal

point process below (Daley & Vere-Jones, 2003).

Let t_i represent timestamps where $t_i < t_{i+1}\ \forall i \in \mathbb{N}^+$. The set S that contains all t_i is said to be the *realization* of the temporal point process. The intensity function is denoted

$$\lambda(t) = \lim_{s \to 0} \frac{\mathbb{E}[N(t+s) - N(t) \mid \mathcal{H}_{t-}]}{s}$$

where t is the current timestamp and s is the timestamp of the event immediately succeeding t. $N(t)$ is the case count at time t where $N(0) = 0$ and $N(t) \geq 0\ \forall t \in S$. Thus, $N(s+t) - N(t)$ is the difference in case count since the previous event. $\mathcal{H}_{t-}$ is the history of the process up to time t, exclusive. More formally, $\mathcal{H}_{t-}$ is the σ-algebra generated by the previously realized points, where σ-algebra is a set A of subsets A_i in $\mathcal{H}$ that satisfy three conditions: $A \in \mathcal{H}$, $A^c \in \mathcal{H}$, the union of all n subsets $A_1, A_2, ..., A_n \in \mathcal{H}$ (Daley & Vere-Jones, 2003). Notice the resemblance with the derivative formula, where $f'(x)$ is the instantaneous rate of change, or derivative, of f at x iff $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}\ \forall x \in Dom(f)$. $\lambda(.)$ is the instantaneous rate at which events are expected to occur at time t, given its prior history $\mathcal{H}_{t-}$ (Park et al., 2020).

Point processes are typically assumed to follow a Poisson distribution. The technical name is *Poisson point process* and is often shortened to *point process* given its conventional popularity. If λ is constant, indicating that events occur at a constant rate, it is considered a *Stationary* or *Homogeneous Poisson Process* (Guhaniyogi & Kang, 2021). Here, $N(t) \sim Poisson(\lambda t)$ and so $E[N(t)] = \lambda t$ by definition of a Poisson distribution, where the Poisson distribution probability density function is formally denoted $P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$ where λ is intensity and k is the number of points (i.e. events). In this particular subtype of Poisson processes, intensity is simply a function of time with a constant rate of occurrence. Time increments are said to be *stationary*. A common application of a Homogeneous Poisson Process is in modeling bus arrivals.

Alternatively, if λ varies with time at a non-constant rate, it is considered a *Non-Stationary* or *Non-Homogeneous Poisson Process*. Here, increments are independent and so event likelihoods can vary with time (Guhaniyogi & Kang, 2021). An example of this is in restaurant customer arrivals, where the rate of arriving customers is dynamic and undulates

throughout the day (i.e. breakfast, lunch, dinner).

3.2 Hawkes Point Process

A Hawkes point process is a method to model self-exciting events (Hawkes, 1971). Most notably, Hawkes models have been used to model events such as earthquakes (Fox et al., 2016), social media interactions (Rizoiu et al., 2017), and terrorist activity (Porter & White, 2012). For example, in Mohler et al. (2011), the onset of one crime was modeled to trigger an initial increase in the likelihood of another crime occurring soon thereafter and with a decaying intensity as time passes. Mathematically, the Hawkes intensity function $\lambda_H(t)$ can be formally expressed as

$$\lambda_H(t) = \mu + \sum_{t_i<t} \varphi(t - t_i)$$

where μ is the exogenous intensity, also known as background or baseline intensity. t is the current time and similarly t_i is the timestamp of each event $i \in \mathbb{N}^+$. $\varphi(.)$ is a triggering kernel function which is some monotonically decreasing function that governs the intensity's decaying behavior between events. Here, I will use an exponential decay function,

$$\varphi(x) = \alpha\beta e^{-\beta x}$$

where α is an inactivity factor and describes the average number of new occurrences triggered by any given event occurrence. β is the exponential decay rate, also known as delay or latency rate. The product of the two is redefined for readability as $\kappa = \alpha\beta$, where κ is referred to as the contagious rate. Thus, the Hawkes model in its entirety is expressed as

$$\lambda_H(t) = \mu + \kappa \sum_{t_i<t} e^{-\beta(t-t_i)}$$

3.3 Neyman-Scott Model

The Neyman-Scott model is a point process model and, as such is suitable for describing clusters of points in time and/or space (Neyman & Scott, 1958). It is a special case of the Cox, or *Doubly Stochastic Poisson Process*, model in which the intensity is stochastic. It was initially developed to describe the interrelated proximity of stars and planets within galaxies. It is a non-causal clustering model and so acts as the baseline model to which we will compare a Hawkes causal clustering model.

In the Neyman-Scott model, there exist parent points and offspring points. Parent points are cluster centers, identified here using Gaussian mixture modeling, a model-based clustering technique in which each component is assumed to independently follow a Gaussian distribution. These points are thought to be hidden and are inferred by the presence of their clustered offspring. Offspring points are then randomly clustered around parents, typically assuming a Poisson distribution. It may be helpful to think of a parent point as being an invisible black hole and offspring points as the surrounding stars and planets. Here, a parent point is an extremely intense disciplinary event in which a high number of students are disciplined at once (e.g. cafeteria food fight), with offspring points being the disciplinary actions taken either before or afterward and for unrelated reasons.

Mathematically, the Neyman-Scott intensity function λ_{NS} can be expressed as

$$\lambda_{NS}(t) = A \sum_{i=1}^{M} C_i$$

where A is the mean cluster size, the average number of offspring classified to each parent. M is the number of clusters, determined using model-based clustering. Here, the expectation-maximization algorithm is used which, in essence, is a blend of k-means and naive Bayes clustering. C describes the offspring clustering process, assumed to follow a Poisson distribution with mean determined by model-based clustering.

3.4 Hypothesis Test

We now perform a hypothesis test formulated by Kresin (2023) and McGovern (2024). It is conventional for the null hypothesis to describe the test's baseline outcome. In the context of testing for causal triggering within student misconduct data, the null hypothesis shall represent the outcome such that student misconduct data are not likely the result of social contagion. The alternative hypothesis shall say that the data are likely the result of social contagion.

Social contagion requires, first, an opportunity for proximal contamination. Only after an initial epoch can subsequent offspring events occur. Hence, evidence for such a phenomenon governing a time series is likely to be exhibited by offspring points densely clustering *after* an initial parent event and not before. The temporal Neyman-Scott model allows for offspring to be placed before and after their respective parents whereas the temporal Hawkes model allows offspring to be placed only after their respective parents. Hence, the Neyman-Scott point process will act as the null model while the Hawkes point process acts as the alternative model. To quantify these hypotheses, we use an information gain statistic in concert with a p-value. This hypothesis test may formally be denoted:

$$\hat{I} = \frac{1}{n}(\log(L_1) - \log(L_0)) = \frac{1}{n} log\left(\frac{L_1}{L_0}\right)$$

$$p = \frac{1}{n}\sum_{i=1}^{n} B_i \quad \text{, where } B_i = \begin{cases} 1 & \text{if } X_i \geq \hat{I} \\ 0 & \text{otherwise} \end{cases}$$

$$H_0 : p \leq \alpha$$

$$H_a : p > \alpha$$

where X_i describes each observed data point. p is the p-value. α is the level of confidence, chosen to be $\alpha = 0.05$. n is the total number of observations. L_0 is the likelihood of the null model (Neyman-Scott). L_1 is the likelihood of the alternative model (Hawkes).

This statistic tracks the magnitude to which information is gained from the null model to the alternative model (Daley & Vere-Jones, 2004). Note, for this test to be viable, it

is necessary for the null model to fit the data relatively well. Otherwise, comparing the alternative model to an ill-fitting null model could give false positives. Evaluation metrics were calculated and placed in Table 4.4 along with comparison plots of residuals as shown in Figures A.5 - A.7.

CHAPTER 4

Results

The following chapter displays results pertaining to tests performed for causal clustering on student misconduct data. Note, these tests were conducted on one high school (High School A), one middle school (Middle School B), and one elementary school (Elementary School C) over four school years: 2016/17, 2017/18, 2018/19, and 2022/23. This was done to allow for intriguing comparable insights. For each test, I compare evaluation metrics, fitted model parameters, test statistics, and other information pertinent to the discussion found in Chapter 5, where I elaborate with more narrative-style writing. Here, results will be shown and objectively interpreted. More subjective inference and claims are reserved for Chapter 5. The structure of this chapter will align with the structure of the procedure specified at the outset of Chapter 3. Many elucidating figures are placed in Appendix A as well.

4.1 Neyman-Scott Fitted Parameters

School	**School Year**	M	A	σ
	2016/17	2	66.5	20.2
High	2017/18	3	70.66	17.73
School A	2018/19	4	64.5	13.4
	2022/23	3	130.66	16.9
	2016/17	6	6.5	5.85
Middle	2017/18	8	158.5	7.4
School B	2018/19	4	367.25	13.6
	2022/23	4	221.75	14.3
	2016/17	2	73	25.75
Elementary	2017/18	3	16.66	16.6
School C	2018/19	1	75	49.46
	2022/23	3	26.66	18.5

Table 4.1: Table displaying fitted parameters of the Neyman-Scott model

As was done in the test procedure, we first fit a Neyman-Scott model to the data using a Gaussian triggering density. M describes the number of clusters, determined using Gaussian mixture modeling and evaluated using Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) scoring metrics, as shown in A.1. A describes the mean cluster size, interpreted as the average number of off-spring assigned to each parent. σ describes the mean of the standard deviations for each cluster. Recall from Section 3.3, a Poisson distribution is used to simulate off-spring around each parent cluster. These parameters are displayed to provide insight into the Neyman-Scott model fit for each school and school year.

4.2 Hawkes Fitted Parameters

School	**School Year**	μ	κ	β
	2016/17	0.3601	0.5084	1.7919
High	2017/18	0.5816	0.4980	0.7953
School A	2018/19	0.8796	0.3854	1.8143
	2022/23	0.6703	0.6949	0.9228
	2016/17	0.1470	0.3947	1.3149
Middle	2017/18	2.6567	0.6101	2.0233
School B	2018/19	3.0464	0.6220	3.0427
	2022/23	2.2973	0.5390	2.3902
	2016/17	0.6326	0.2719	1.2286
Elementary	2017/18	0.1896	0.3839	1.8459
School C	2018/19	0.3889	0.0344	0.0635
	2022/23	0.3043	0.3329	3.4933

Table 4.2: Table displaying fitted parameters of the Hawkes model

Next, a Hawkes model was fit to the data using the Expectation-Maximization algorithm. μ describes the background intensity, interpreted as the baseline intensity. A large μ value indicates an overall high degree of student misconduct in a given school year. Notice, the largest μ value in High School A occurs in the 2018/19 school year. The μ values in Middle School B are extremely high over the span of three years beginning in the 2017/18 school year and continuing to the 2022/23 school year. In Elementary School C, the highest μ value occurs in the 2016/17 school year. κ describes the rate of contagion. A large κ value indicates a highly contagious event. Notice the high κ values for High School A in 2022/23, Middle School B in 2017/18 and 2018/19, and Elementary School C in 2022/23. β is the exponential decay rate, governing the speed at which intensity decreases. A large β indicates a fast decrease in intensity and indicates a low frequency of events in the given school year.

4.3 Hypothesis Test

School	School Year	P-Value of Hypothesis Test
	2016/17	0
High	2017/18	0
School A	2018/19	0
	2022/23	0.938
	2016/17	0
Middle	2017/18	1
School B	2018/19	1
	2022/23	1
	2016/17	0
Elementary	2017/18	0
School C	2018/19	1
	2022/23	0

Table 4.3: Table displaying hypothesis tests' p-values

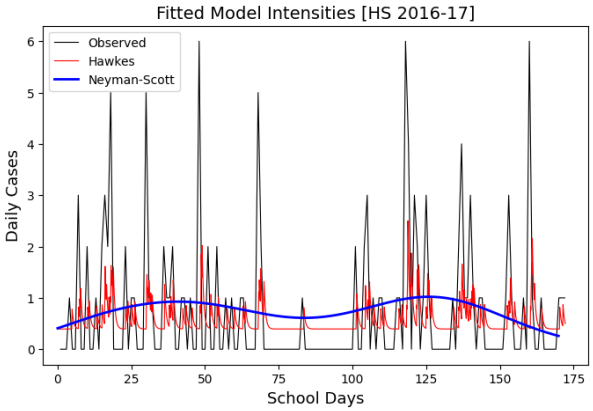
Finally, we performed hypothesis tests comparing the distribution of 500 simulated information gain statistics to that of the observed value for each school and school year. A p-value is calculated for each to determine whether or not to reject the null hypothesis. Here, the p-value is calculated as the proportion of simulated information gain statistics that are as or more extreme than the observed information gain statistic. Notice the extreme p-values; most of which are either near 0 or near 1, indicating extreme differences in each model's fit. The tests where the null hypothesis failed to be rejected include: High School A in 2022/23, Middle School B in 2017/18, 2018/19, 2022/23, and Elementary School C in 2018/19.

4.4 Evaluation Metrics

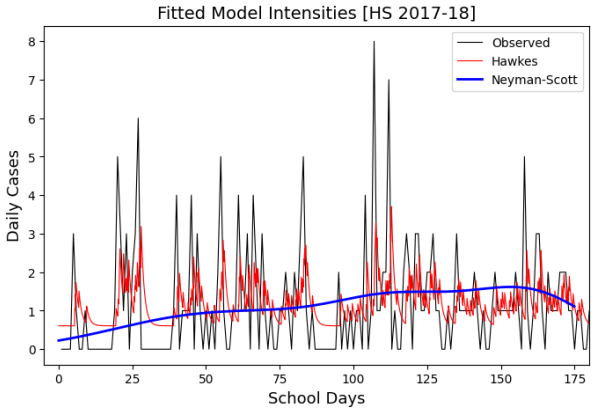
School	School Year	Hawkes RMSE	Neyman-Scott RMSE
	2016/17	1.28	1.26
High	2017/18	1.47	1.39
School A	2018/19	1.73	1.66
	2022/23	2.36	2.36
	2016/17	0.6	0.6
Middle	2017/18	4.97	4.86
School B	2018/19	6.04	5.29
	2022/23	4.14	3.87
	2016/17	1.14	1.13
Elementary	2017/18	0.87	0.75
School C	2018/19	0.68	0.64
	2022/23	1.01	1.03

Table 4.4: Table displaying Root Mean Squared Error (RMSE) of fitted Hawkes and Neyman-Scott models

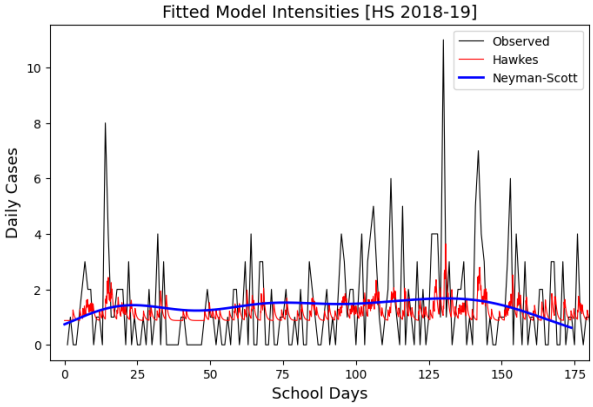
To ensure the null model (Neyman-Scott) fits the data well enough to act as a viable baseline model, we calculate the Root Mean Squared Error (RMSE) evaluation metric. A low RMSE indicates an accurate model. For example, in the fitted Hawkes model for High School A in 2016/17, the RMSE states that the prediction deviated by an average intensity of 1.28 daily events (i.e. disciplinary actions). Notice, the Neyman-Scott model's RMSE is less than or equal to that of the associated Hawkes model in all but one test (Elementary School C, 2022/23). This indicates the Neyman-Scott model is indeed a viable model for comparison. Plots comparing the fitted models to observed data are shown below.



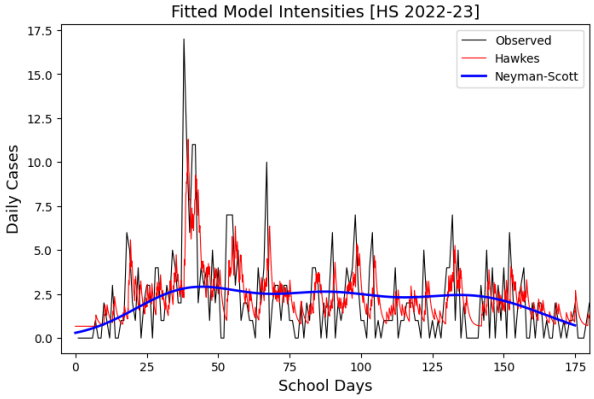
(a) School Year 2016/17



(b) School Year 2017/18



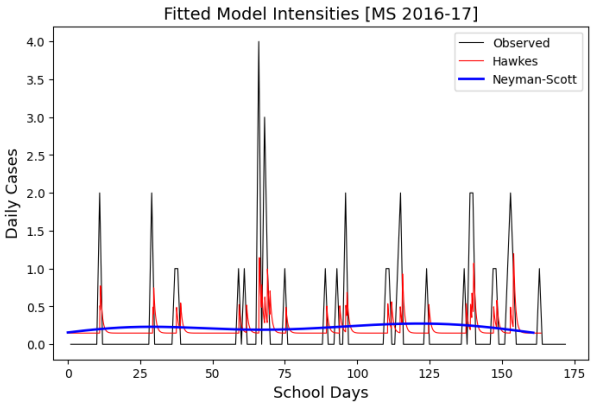
(c) School Year 2018/19



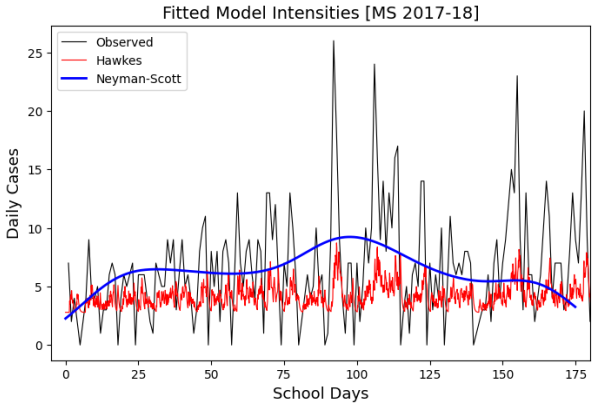
(d) School Year 2022/23

Figure 4.1: Fitted Neyman-Scott and Hawkes Models - High School

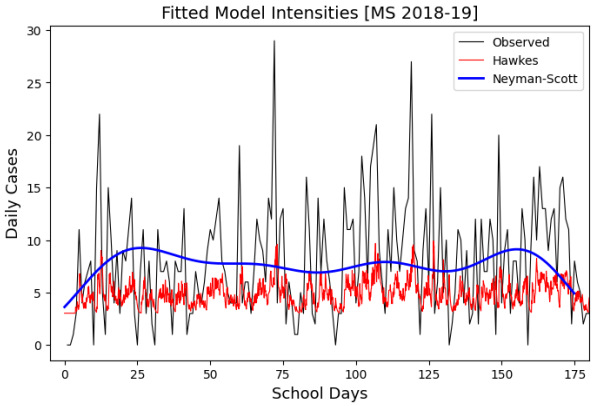
Figure 4.1 compares the fitted models with that of the observed data for high school. One can clearly see the Newman-Scott model (blue) fitting parent points to the most extreme intensities observed throughout each school year, with local maxima being attracted to the largest intensities.



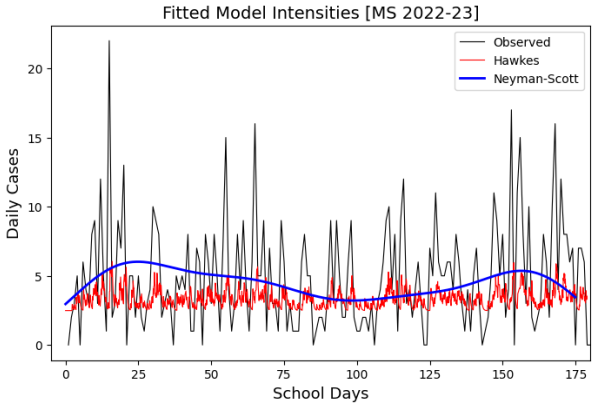
(a) School Year 2016/17



(b) School Year 2017/18



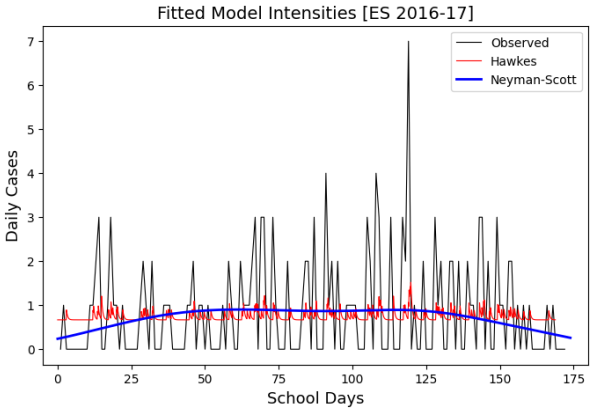
(c) School Year 2018/19



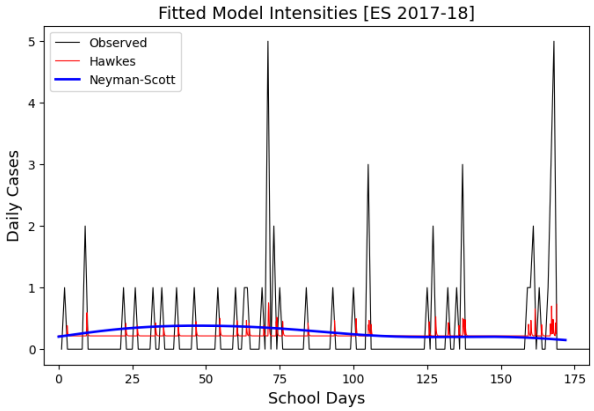
(d) School Year 2022/23

Figure 4.2: Fitted Neyman-Scott and Hawkes Models - Middle School

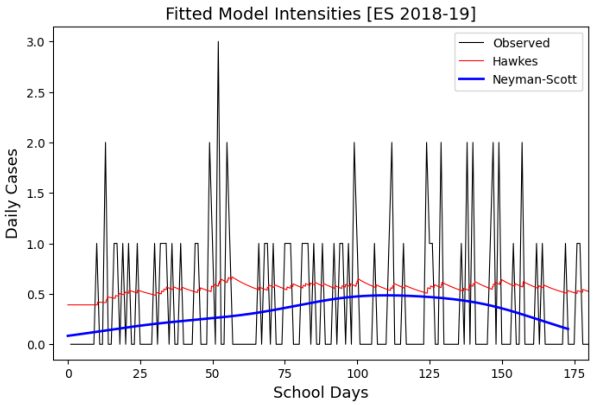
Figure 4.2 compares the fitted models with that of the observed data for middle school. Notice the low frequency of events in 2016/17 compared to the high frequency of events in other years. As discussed earlier and shown in Table 4.3, the school years with a high frequency of events fail to reject the null hypothesis compared to years with a low number of events. In essence, the excitation modeled by Hawkes models (red) is diluted by the high frequency of events (black); as a result, hypothesis tests are less able to conclude the presence of significant differences between the fitted Hawkes and Neyman-Scott models.



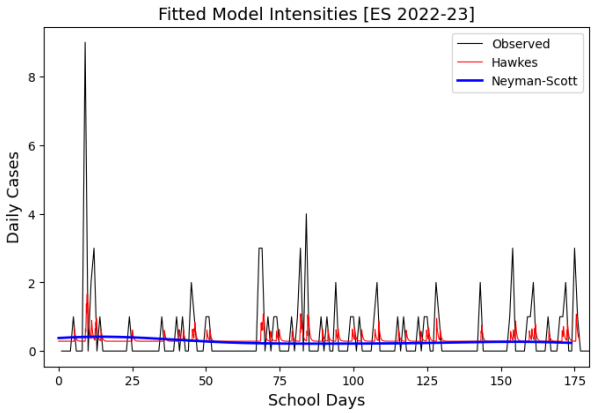
(a) School Year 2016/17



(b) School Year 2017/18



(c) School Year 2018/19



(d) School Year 2022/23

Figure 4.3: Fitted Neyman-Scott and Hawkes Models - Elementary School

Figure 4.3 compares the fitted models with that of the observed data for elementary school. Similar behaviors are shown here, in elementary school, as described and portrayed previously in middle and high school. Similarly to before, the frequency of events appears to govern the difference between the Hawkes and Neyman-Scott model fits.

4.5 Case Count Proportions

Table 4.5 displays the annual case count and calculated proportion of disciplinary actions that occurred for each school. Proportion is calculated by taking the ratio of annual case count and enrollment. This is done to standardize the results for viable comparison across

School	School Year	Case Count	Proportion
	2016/17	125	0.1835
High	2017/18	212	0.3122
School A	2018/19	258	0.3583
	2022/23	392	0.4768*
	2016/17	39	0.0615
Middle	2017/18	1268	1.829*
School B	2018/19	1469	2.069*
	2022/23	887	1.261*
	2016/17	146	0.2588
Elementary	2017/18	50	0.0877
School C	2018/19	75	0.3*
	2022/23	80	0.27

Table 4.5: Table displaying annual case count by school

schools and years. Notice, for all tests, the case proportion is highest for its respective school and across school years when the associated p-value is less than 0.05 (i.e. the hypothesis is failed to be rejected), shown in Table 4.3. The asterisk (*) indicates the tests in which the null hypothesis failed to be rejected.

CHAPTER 5

Discussion

In this chapter, I provide further insight, inference, and narrative pertaining to the results of the tests for causal clustering on the event of student misconduct. As shown in Table 4.3, tests varied. Out of the twelve tests performed, seven were able to reject the null hypothesis. Interestingly, both event frequency and event magnitude seem to hold strong governance over their associated test conclusion.

Considering the case count proportions displayed in Table 4.5, the tests in which the p-value exceeds the significance level α directly corresponds to those with high case count proportions. Recall, case count proportion is calculated as the ratio of annual cumulative case count and total enrollment. This indicates that the test used to determine causal triggering is governed, at least in part, by the frequency of event occurrences. Here, a higher frequency of occurrences corresponds to higher p-values, and thus, lower confidence that the data is the product of causal triggering. To illustrate this, let us consider the plots displaying fitted Hawkes and Neyman-Scott models found in Figures A.2 - A.4.

First we consider the hypothesis tests for High School A in Figure A.2. For the first three school years tested, there is statistically significant evidence of causal triggering. However, we fail to reject the null hypothesis in the final school year, 2022/23. I want to bring your attention to the varying y-axis limits across school years. For example, intensity ranges from 0-17 in the 2022/23 school year whereas, in 2016/17, the range is 0-6. By inspection, it is immediately clear that the Hawkes model offers a more flexible fit than that of the Neyman-Scott model. Because of its agility, the Hawkes model is better able to register extreme intensities compared to its Neyman-Scott counterpart. This finding aligns with those found in McGovern (2024). As mentioned previously, there is also a higher frequency of occurrences

in 2022/23 compared to earlier years (see Table 4.5 and Figure 4.1). Hence, for the test that failed to reject the null hypothesis, the school year consisted of both a high event frequency and a high event magnitude, relative to that of all school years studied in High School A.

Next we consider the hypothesis tests for Middle School B in Figure A.3. Here, three out of four tests failed to reject the null hypothesis, school years: 2017/18, 2018/19, and 2022/23. Similar to what was discussed in High School A, the event magnitudes for these school years are much higher than in 2016/17, which is the only school year in Middle School C that showed evidence of causal triggering. In 2016/17, the maximum case count was 4, whereas in later years, the maximum case count reached upwards of 30 daily occurrences. Moreover, the frequency of events is much higher in 2017/18, 2018/19, and 2022/23 than it is in 2016/17. This can be seen visually in Figure 4.2 and numerically in Table 4.5. In Figure 4.2, you can see the Hawkes model struggle to respond to such drastic fluctuations in cases. Like in the tests performed for High School A, the high frequency of occurrences coupled with extreme fluctuations in daily occurrences exacerbates the failure of rejecting the null hypothesis.

Lastly, we consider the hypothesis tests for Elementary School C in Figure A.4. Three out of four tests rejected the null hypothesis, showing statistically significant evidence of casual triggering in school years: 2016/17, 2017/18, and 2022/23. As is true in High School A and Middle School B, the test that failed to reject the null hypothesis is for the school year that had a noticeable uptick in both case frequency and proportion (see Table 4.5). By visually inspecting the model plots in Figure 4.3, the Hawkes model in 2018/19 learned a low decay rate ($\beta = 0.0635$) due to a high frequency of occurrences; thus, it is not accurate in modeling days with low case count (see the plot of residuals in Figure A.7). As a result, the Hawkes model fit for 2018/19 school year fails to outperform the more conservative Neyman-Scott fit.

CHAPTER 6

Conclusion

6.1 Overview

This paper concludes with a brief summary of notable claims and insights. First, the dynamic behavior of student misconduct was studied broadly within and across school years. Descriptive plots were used to evince the relatively flat trend of student misconduct rates from 2016 to 2023. Disciplinary actions experienced dormant periods at the beginning of the school year, around winter break, around spring break, and near the end of the school year. Moreover, disciplinary actions appear to be significantly higher before the pandemic compared to after. This may, however, be explained by an increase in absenteeism.

After performing exploratory data analysis, hypothesis tests were conducted to test for causal triggering. Tests varied across schools and school years; however, intriguing relationships were revealed in the posterior analysis. Specifically, both frequency and magnitude govern whether or not the test for causal triggering can reject the null hypothesis. High frequency and extreme fluctuations appear to inhibit the test's ability to reject the null hypothesis.

6.2 Future Research

Many future areas of statistical research hold great potential within the field of education and should be addressed in subsequent research. Here, I will focus on research efforts involving student misconduct, with an emphasis on statistical modeling techniques.

First, applying the test for causal triggering, as performed in this paper, to a more promi-

nent municipality could yield a more robust understanding of the dynamic interactive nature of adolescent misbehavior and adult disciplinary enforcement. Moreover, studying a region with a high density of schools would provide a more suitable context for spatiotemporal point processes to be applied, similar to what has been done with criminal activity and earthquakes (Mohler et al., 2011; Fox et al., 2016). From this, the formation of heat maps could provide further insight into the social ecosystem of the region, allowing for relationships to be studied that connect school-level factors to more global social factors. For example, is there a relationship between criminal activity and student misbehavior? Does the occurrence of a violent crime such as gang violence excite the behavior of student misconduct? Are schools with high rates of student misconduct located in areas with high rates of criminal activity? Researching causal discovery within the context of education offers many opportunities to learn more about the nature of social dynamics within a region. In particular, investigating causal discovery between school shootings and student misconduct could be especially insightful.

Second, involving data on the type of student misbehavior could yield insightful findings. As noted in Section 2.1, this information was omitted to protect student privacy and so all misbehavior types are treated equally. With this data, intensity could be the rank of a disciplinary event wherein the rank reflects the act's severity (e.g. 1 = verbal warning, 2 = in-school suspension, 3 = out-of-school suspension, 4 = expulsion). Defining rankings will depend on the data available and the school ecosystem. With this information, one could compare the dynamic nature of student misbehavior across types. Do expulsions cause a higher likelihood spike than suspensions? What is the most contagious type of misbehavior? Does enforcing cell phone misuse effectively thwart its spread? Gaining access to more detailed data could help answer such probing questions. Moreover, involving data on school policy, school climate, and type of school (i.e charter, private, public), student misbehavior could be analyzed and modeled to the extreme level of detail that has been achieved, for example, in modeling infectious diseases in the field of epidemiology (Park et al., 2020).

Third, since the COVID-19 pandemic, student absenteeism has risen at alarming rates (Malkus, 2024). Future research should be done to study the interactive relationship between

student misconduct and student absenteeism, both of which falling into the broader category of adolescent behavioral dynamics.

Fourth, students spend the majority of their time within the classroom and so contagion is mainly contained within the classroom with little dissemination elsewhere. This constrains productivity and may lead to underestimation, predominantly for classes with only a small number of students. Formulating a multi-level model partitioned by class size could provide additional insight.

Lastly, the hypothesis test used to determine causal triggering is inhibited by high frequencies and extreme fluctuations. More research should be done to study the influence of these inhibiting factors and formulate strategies to mitigate their adverse effects.

APPENDIX A

Appendix

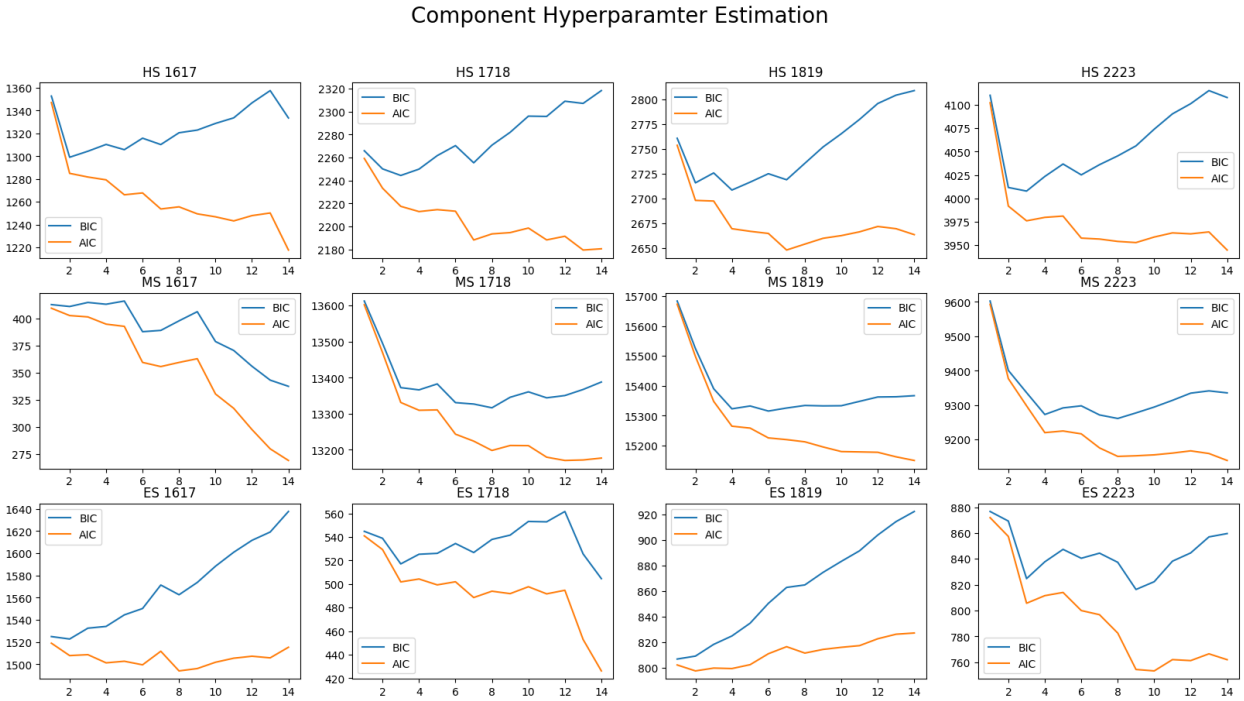
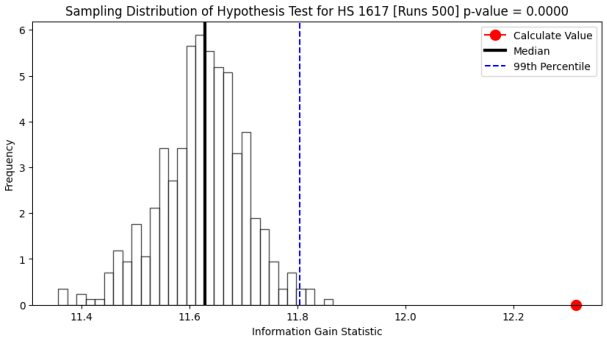
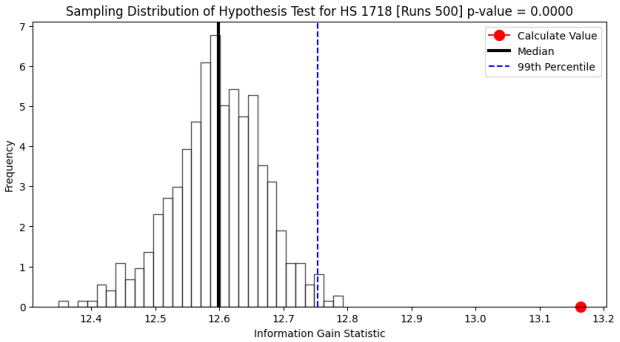


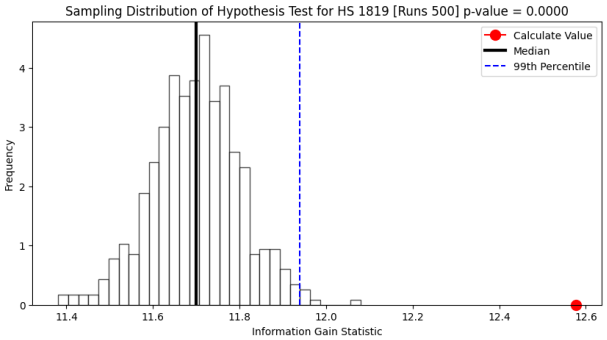
Figure A.1: BIC and AIC metrics determined using Gaussian mixture modeling



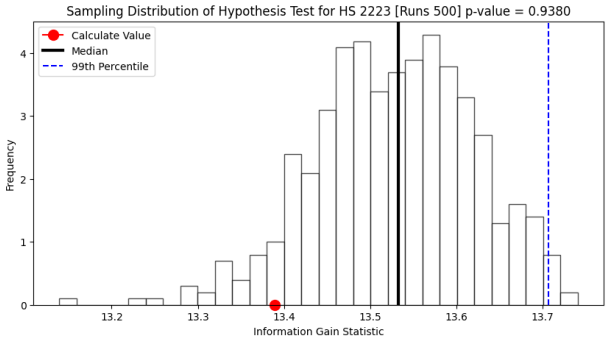
(a) School Year 2016/17



(b) School Year 2017/18

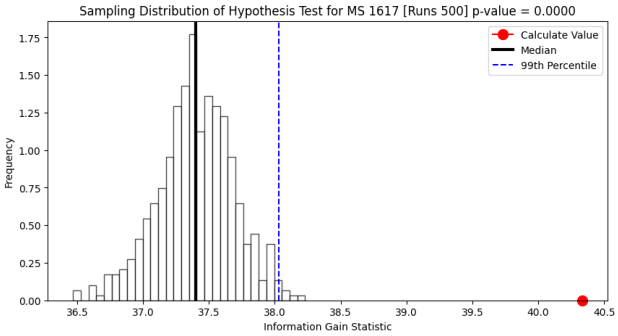


(c) School Year 2018/19

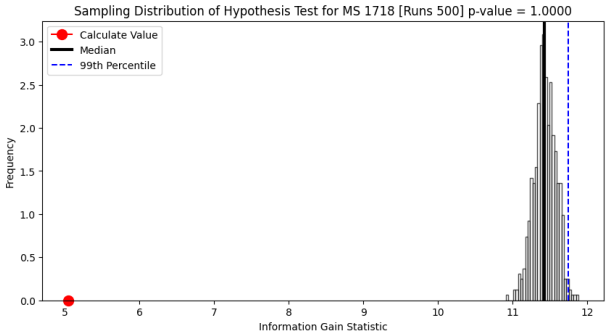


(d) School Year 2022/23

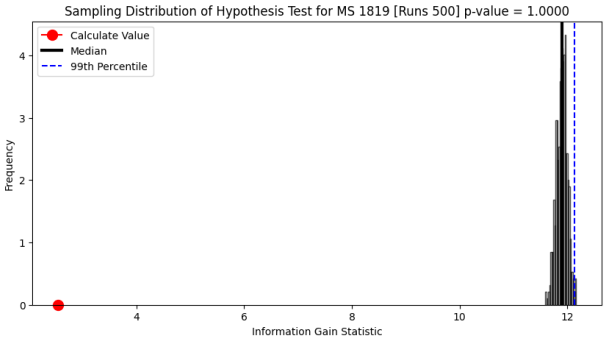
Figure A.2: Information Gain Statistic Plots - High School



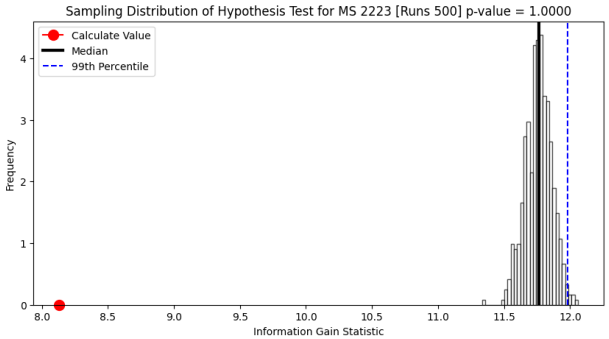
(a) School Year 2016/17



(b) School Year 2017/18

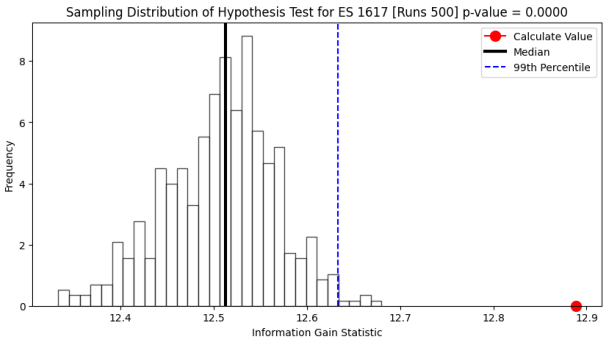


(c) School Year 2018/19

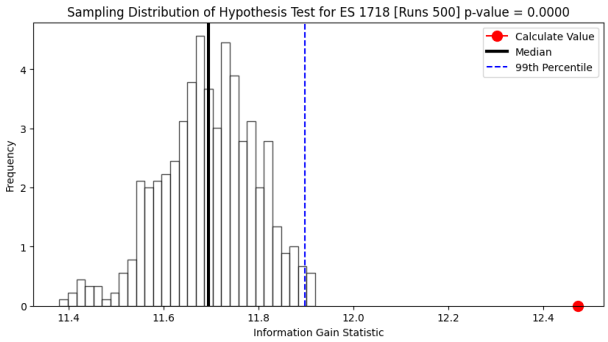


(d) School Year 2022/23

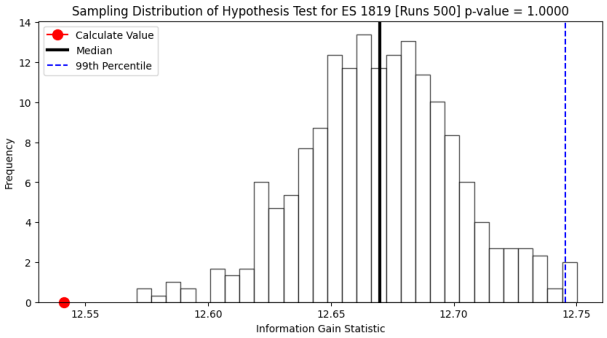
Figure A.3: Information Gain Statistic Plots - Middle School



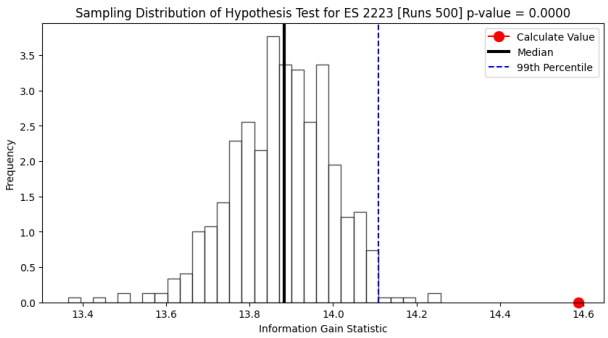
(a) School Year 2016/17



(b) School Year 2017/18

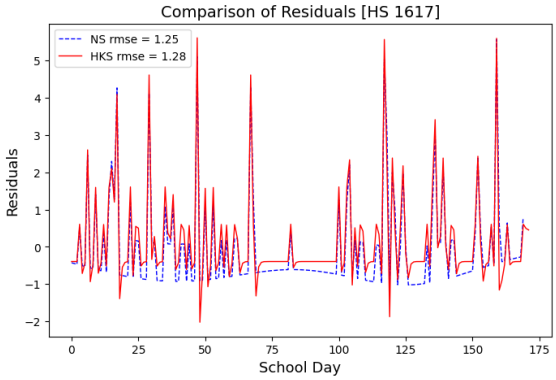


(c) School Year 2018/19

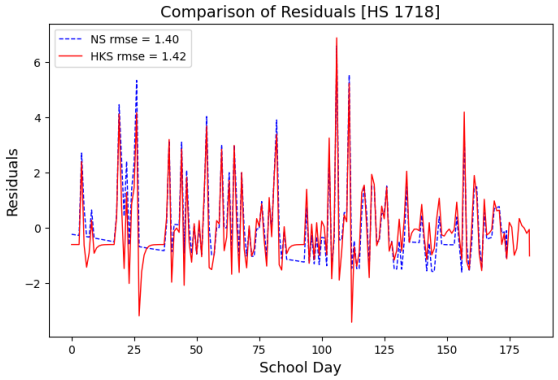


(d) School Year 2022/23

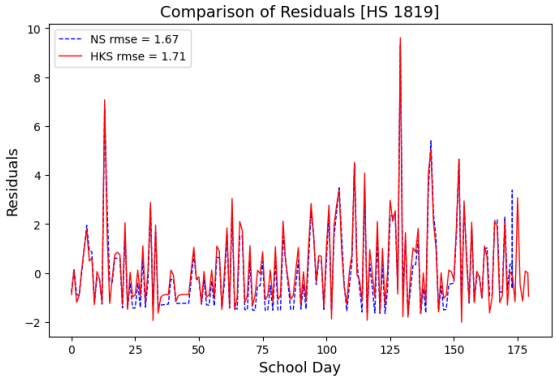
Figure A.4: Information Gain Statistic Plots - Elementary School



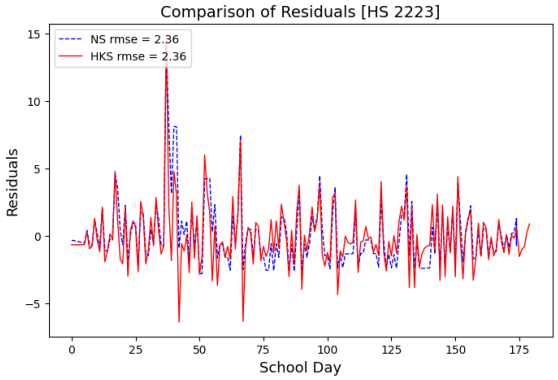
(a) School Year 2016/17



(b) School Year 2017/18

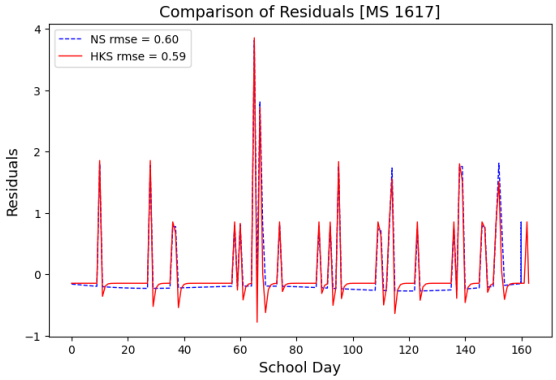


(c) School Year 2018/19

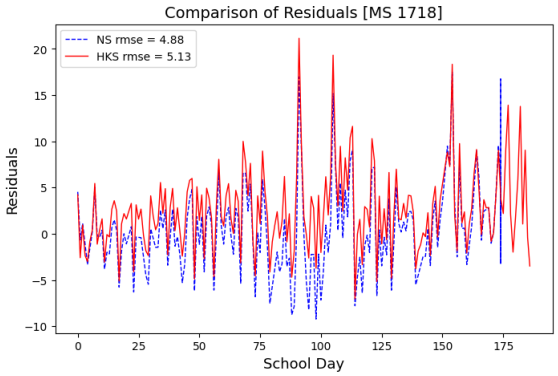


(d) School Year 2022/23

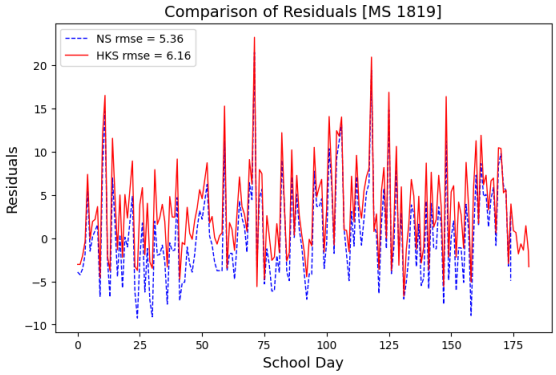
Figure A.5: Comparison of Residuals Plots High School



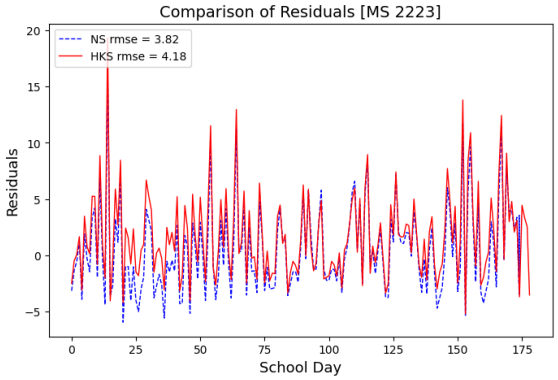
(a) School Year 2016/17



(b) School Year 2017/18

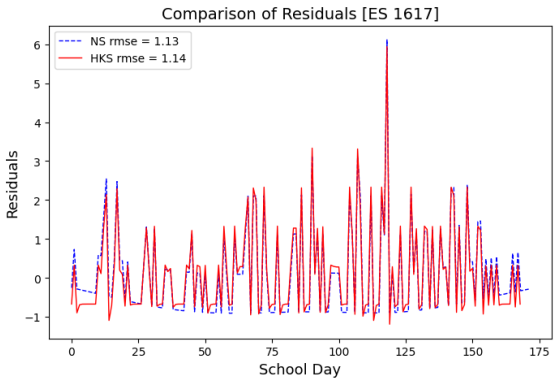


(c) School Year 2018/19

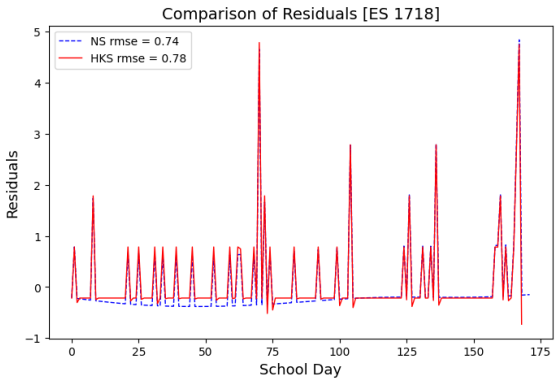


(d) School Year 2022/23

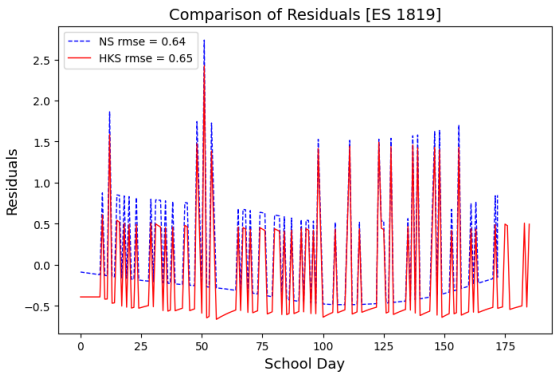
Figure A.6: Comparison of Residuals Plots Middle School



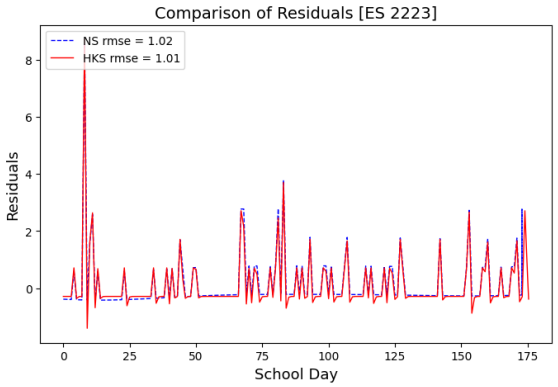
(a) School Year 2016/17



(b) School Year 2017/18



(c) School Year 2018/19



(d) School Year 2022/23

Figure A.7: Comparison of Residuals Plots Elementary School

BIBLIOGRAPHY

Aldric, A. (2024). *Average SAT Scores Over Time: 1972–2023*. https://blog.prepscholar.com/average-sat-scores-over-time (accessed: 05.17.2024).

Daley, D. J., & Vere-Jones, D. (2004). Scoring Probability Forecasts for Point Processes: The Entropy Score and Information Gain. *Journal of Applied Probability*, *41*, 297–312. https://www.jstor.org/stable/3215984

Daley, D., & Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes: Volume 1: Elementary Theory and Methods, Second Edition*. Springer.

Fabes, R., Quick, M., Catherine, E., & Musgrave, A. (2021). Exclusionary discipline in U.S. public schools: A comparative examination of use in Pre-Kindergarten and K-12 grades. *Educational Studies*, *50*(3), 299–316. https://doi.org/10.1080/03055698.2021.1941782

Fox, E., Schoenberg, F., & Gordon, J. (2016). Spatially Inhomogeneous Background Rate Estimators And Uncertainty Quantification For Nonparametric Hawkes Point Process Models of Earthquake Occurrences. *The Annals of Applied Statistics*, *10*(3), 1725–1756. https://www.jstor.org/stable/43956899

Franken, P. (1982). Point Process Method In Queueing Theory. In: Disney, R.L., Ott, T.J. (eds). *Applied Probability-Computer Science: The Interface Volume 1. Progress in Computer Science, vol 2. Birkhäuser Boston*. https://doi.org/10.1007/978-1-4612-5791-2_6

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424–438. https://doi.org/10.2307/1912791

Guhaniyogi, R., & Kang, J. (2021). *Stochastic Process*. https://bookdown.org/jkang37/stochastic-process-lecture-notes/ (accessed: 05.02.2024).

Hanson, M. (2023). *U.S. Public Education Spending Statistics*. https://educationdata.org/public-education-spending-statistics#:~:text=Public%20education%20spending%20in%20the,fund%20K%2D12%20public%20education (accessed: 05.15.2024).

Hawkes, A. G. (1971). Spectra of Some Self-Exciting and Mutually Exciting Point Processes. *Biometrika*, *58*(1), 83–90. https://doi.org/10.2307/2334319

Kresin, C. J. (2023). Applications and Properties of Point Processes. UCLA. ProQuest ID: Kresin_ucla_0031D_21585. Merritt ID: ark:/13030/m5p638tv. https://escholarship.org/uc/item/1997h647

Malkus, N. (2024). *Long COVID for Public Schools: Chronic Absenteeism Before and After The Pandemic* (tech. rep.). American Enterprise Institute. https://www.aei.org/research-products/report/long-covid-for-public-schools-chronic-absenteeism-before-and-after-the-pandemic/ (accessed: 05.10.2024).

McGovern, I. (2024). Causality in Point Processes. UCLA. ProQuest ID: McGovern_ucla_0031D_22955. Merritt ID: ark:/13030/m5bh3wrd. https://escholarship.org/uc/item/5x56b2fk

Mohler, G., Short, M., Brantingham, P., Schoenberg, F., & Tita, G. (2011). Self-Exciting Point Process Modeling of Crime. *Journal of the American Statistical Association*, *106*(493), 100–108. https://doi.org/10.1198/jasa.2011.ap09546

National Assessment of Educational Progress. (2024). *NAEP Long-Term Trend Assessment Results: Reading and Mathematics*. https://www.nationsreportcard.gov/ltt/?age=13 (accessed: 05.16.2024).

Neyman, J., & Scott, E. L. (1958). Statistical Approach to Problems of Cosmology. *Journal of the Royal Statistical Society*, *20*(1), 1–43. https://doi.org/10.1111/j.2517-6161.1958.tb00272.x

Park, J., Chaffee, A. W., Harrigan, R. J., & Schoenberg, F. P. (2020). A non-parametric Hawkes model of the spread of Ebola in West Africa. *Journal of Applied Statistics*, *49*(3), 621–637. https://doi.org/10.1080/02664763.2020.1825646

Porter, M. D., & White, G. (2012). Self-Exciting Hurdle Models For Terrorist Activity. *The Annals of Applied Statistics*, *6*(1), 106–124. https://doi.org/10.1214/11-AOAS513

Rizoiu, M.-A., Lee, Y., Mishra, S., & Xie, L. (2017). A Tutorial on Hawkes Processes for Events in Social Media. arxiv. https://arxiv.org/pdf/1708.06401 (accessed: 05.07.2024).

Schoenberg, F. (2011). *Introduction to Point Processes*. Wiley Encyclopedia of Operations Research; Management Science. https://doi.org/10.1002/9780470400531.eorms0425

Solow, D. (2013). *How to Read and Do Proofs: An Introduction to Mathematical Thought Processes, 6th Edition*. Wiley.

United Nations Educational, Scientific and Cultural Organization. (2023). *What you need to know about the right to education*. https://www.unesco.org/en/articles/what-you-need-know-about-right-education#:~:text=About%20258%20million%20children%20and,youth%20lack%20basic%20literacy%20skills (accessed: 05.15.2024).

Widder, D. (1989). *Advanced Calculus, 2nd Edition*. Dover Publications.

Zhuang, J., Ogata, Y., & Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*. https://doi.org/https://doi.org/10.1198/016214502760046925