Modeling News Data Flows using Multivariate Hawkes Processes

ERIK ALPSTEN
Modeling News Data Flows using Multivariate Hawkes Processes

ERIK ALPSTEN

Degree Projects in Mathematical Statistics (30 ECTS credits)
Degree Programme in Applied and Computational Mathematics (120 credits)
KTH Royal Institute of Technology year 2018
Supervisor at Lynx Asset Management: Martin Rehn
Supervisor at KTH: Boualem Djehiche
Examiner at KTH: Boualem Djehiche
Modeling News Data Flows using Multivariate Hawkes Processes

Abstract

This thesis presents a multivariate Hawkes process approach to model flows of news data. The data is divided into classes based on the news’ content and sentiment levels, such that each class contains a homogeneous type of observations. The arrival times of news in each class are related to a unique element in the multivariate Hawkes process. Given this framework, the massive and complex flow of information is given a more compact representation that describes the excitation connections between news classes, which in turn can be used to better predict the future flow of news data. Such a model has potential applications in areas such as finance and security. This thesis focuses especially on the different bucket sizes used in the discretization of the time scale as well as the differences in results that these imply. The study uses aggregated news data provided by RavenPack and software implementations are written in Python using the TensorFlow package.

For the cases with larger bucket sizes and datasets containing a larger number of observations, the results suggest that the Hawkes models give a better fit to training data than the Poisson model alternatives. The Poisson models tend to give better performance when models trained on historic data are tested on subsequent data flows. Moreover, the connections between news classes are given to vary significantly depending on the underlying datasets. The results indicate that lack of observations in certain news classes lead to over-fitting in the training of the Hawkes models and that the model ought to be extended to take into account the deterministic and periodic behaviors of the news data flows.
Modellering av Nyhetsdataflöden med Multivariata Hawkesprocesser

Sammanfattning


För testerna med större tidsskalor och dataset som innehåller större mängd observationer ger resultaten att hawkesmodellerna anpassas bättre till träningsdata än de enklare poissonmodellerna. Dock tenderar poissonmodellerna ge bättre prestanda när modellerna som tränats på historiska data sedan testas på efterföljande nyhetsdataflöden. Dessutom fás att kopplingarna mellan nyhetsklasserna varierar avsevärt beroende på underliggande dataset. Resultaten tyder på att bristen på observationer i vissa nyhetsgrupper leder till överpassning i träningen av hawkesmodellerna och att modellen bör utvidgas för att bättre ta hänsyn till de fenomen i nyhetsdataflödet som är deterministiska och periodiska.
Acknowledgements

To begin with, I would like to thank my KTH supervisor Boualem Djehiche for guidance and useful advice. I would also like to express my gratitude to Lynx Asset Management for giving me the opportunity to conduct this project. In particular, I would like to thank my company supervisors Martin Rehn, Ola Backman, Jing Fu and Tobias Rydén for always providing insightful ideas and inspiring me to be the best I can ever be. I would also like to thank my thesis brother-in-arms Carl Brishammar, with whom I have shared this experience.

Last but not least, I would like to thank my family and friends who have always been there. You will always remain the most important element in life.

Stockholm, May 2018

Erik Alpsten
Disclaimer

This thesis was conducted as part of a collaboration between me, Erik Alpsten from KTH Royal Institute of Technology, and Carl Brishammar from Lund University. The two of us cooperated extensively throughout the projects and shared data and mathematical models as well as methods and software implementations. However, the thesis projects were counted as separate and we therefore submitted individual reports to our respective home universities. Thus, each report focuses on different topics of analysis and results. The common report sections however, such as introduction, data, mathematical background, implementation and methods, contain content shared between the projects with only smaller variations between the two reports.
## Contents

1 Introduction .......................... 1
   1.1 Related Work .......................... 2
   1.2 Scope, Objectives & Limitations ........... 3
   1.3 Report Outline .......................... 4

2 News Data .......................... 5
   2.1 Aggregated News Data .................... 5
   2.2 The RavenPack Data ....................... 6
       2.2.1 Dataset Overview ..................... 6
       2.2.2 Visualizations of Important Fields ....... 8

3 Mathematical Background .............. 13
   3.1 Stochastic Processes ..................... 13
       3.1.1 Basic Stochastic Processes .......... 13
       3.1.2 The Hawkes Process ................ 15
   3.2 Modeling News Data ...................... 20
       3.2.1 Distinct Classes ..................... 21
       3.2.2 Overlapping Classes ................ 23
   3.3 Optimization & Parameter Estimation ...... 25
       3.3.1 Gradient Descent ................... 25
       3.3.2 ADAM .............................. 26
   3.4 Statistical Model Evaluation ............. 27

4 Methods ........................... 29
   4.1 Bucketing ................................ 29
   4.2 Implementation of Distinct Classes Model .... 31
   4.3 Setup ................................ 34
       4.3.1 Construction of Classes .......... 34
       4.3.2 Spatial Distribution ............... 35
       4.3.3 Inhomogeneous Extension .......... 36
   4.4 Settings & Test Scenarios ............... 37
5 Results

5.1 Bucket Size: Day ........................................ 39
   5.1.1 Homogeneous Model .............................. 39
   5.1.2 Inhomogeneous Model ......................... 44
   5.1.3 Algorithm Convergence ....................... 49

5.2 Bucket Size: Hour ..................................... 53

5.3 Bucket Size: Minute ................................. 58

5.4 Connections Between Classes ..................... 63

6 Discussion ................................................... 68

6.1 Discussion of Results ............................... 68

6.2 Returning to Scope & Objectives ................ 70

6.3 Future Work .............................................. 71

7 Conclusions .................................................. 73

References .................................................... 74

Appendices ..................................................... 76

A List of IDs and News Classes .......................... 76

B Observations per News Class for Yearly Datasets ........................................ 79

C Observations per News Class for Monthly Datasets ......................................... 82

D Observations per News Class for Daily Datasets ............................................. 85
Chapter 1

Introduction

In some applications it is of importance to quickly gather information and react to events around the world. This is the case in areas such as security, health and finance. However, it can often be both costly and time-consuming to gather information from primary sources. In such a case, it may be easier or even necessary to rely on reports from secondary sources such as news articles. This opens up the question if it is possible to automatically extract relevant information from news as well as how to model and react to the news data flows. Historically, news has been a widespread and important way of communicating information. This has been done through numerous channels of communication, such as radio, television, newspapers and online news sources to only name a few. Even though news from these sources are different in structure and availability, most news types share some important properties that are central to this study.

To begin with, an important characteristic of news in general is the clustering of data about specific topics around certain time points. For instance, if the news category of interest is earthquakes, it is unlikely that there will be a uniform distribution of earthquake news over longer time periods. Instead, it is more likely to observe clusters of news about earthquakes around specific time points. Generally, this occur due to real-world events that causes the increase and cluster of news about the specific topic. That is, if an earthquake has just occurred there will most likely be a lot of news coverage about the event within the near future. In a similar way, an absence of news data points about the topic can often be explained by the fact there have not been any relating events for some time, for example if the latest earthquake took place several months or years earlier.

With this insight about the structure of news data flows, mathematical methods that take the clustering characteristics into account could be suitable candidates to model the flow of news data. Here, the times of occurrence
and the extracted contents of the news data are useful attributes that can be used to model a stochastic environment that represents the flow of news data. However, to incorporate the clustering characteristic introduced above it is of interest to use a model that can adapt to this property. One such model is the Hawkes process model.

In short, the Hawkes process is a generalization of the Poisson process in the sense that its intensity depends on the history of the process. More specifically, the intensity function is self-exciting, which means that observing points from the process increases the intensity in the near future, thus providing a model for the clustering phenomenon. Formal mathematical definitions of these concepts are presented in Chapter 3.

1.1 Related Work

Analysis and prediction of news data have gained increased interest in recent years. This is partially due to the attention from the financial sector, but also from news providers, social media channels and other organizations looking to optimize their user experience, marketing efforts and other operations. The spectrum of analysis is rather wide and involves many different stages, e.g. natural language processing to interpret the text, data mining to handle large sets of information as well as a range of statistical methods to model data flows. The paragraphs below presents a selection of related works that is relevant to the background of this thesis project.

The topic of news analytics as a method in the financial sector is discussed extensively in the book "The handbook of news analytics in finance" [1], which presents several techniques in handling news data as well as its potentials and risks in predicting financial assets. Similarly, the articles "News vs. sentiment: predicting stock returns from news stories" and "Stock price prediction using financial news articles" [2, 3] both deal with prediction of stock prices using news data. The first article presents a support vector machine approach using features extracted from financial news articles and historic stock prices, whereas the second article examines the prediction accuracy of neural networks for stock returns. Finally, the article "Applications of a multivariate Hawkes process to joint modeling of sentiment and market return events" [4] explores the use of point processes and Hawkes processes to model events in financial markets. More specifically, the study analyzes how positive and negative sentiments in news events connect to positive and negative returns in the context of multivariate Hawkes processes.

Another related area is that of modeling and prediction of events on social media, e.g. how content goes viral and spreads on different channels as well as how it can be used to predict events outside social media platforms. The
article "Predicting the future with social media" [5] uses data from Twitter
to forecast the revenues of box-office movies. In addition, the article "A
survey of prediction using social media" [6] discusses several topics within
the subject, such as marketing, information validity and prediction of election
outcomes. Lastly, the article "A tutorial on Hawkes processes for events in
social media" [7] provides an introduction to the concept of Hawkes processes
and the self-exciting properties, with a focus on social media events.

In addition to the financial applications introduced above, the Hawkes pro-
cess has been used in for example earthquake forecasting as well as modeling
epidemic outbreaks. The common characteristic in these areas is the self-
exciting property. For instance, for epidemic diseases it may be reasonable
to suggest that observing a case of the disease in a certain area will increase
the risk, i.e. the intensity, to observe more cases in that region within the
near future, thus making the Hawkes process model a suitable candidate.
One such study is "A recursive point process model for infectious diseases"
[8], which uses Hawkes process as well as another type of point process to
model measles occurrences between 1906 and 1956. Another relevant article
is "Assessment of point process models for earthquake forecasting" [9], which
reviews the Hawkes process among other model alternatives for earthquake
forecasting.

1.2 Scope, Objectives & Limitations

The general objective of this thesis project is to build and evaluate a multi-
variate Hawkes process model for the flow of news data. More specifically,
given the sets of aggregated news data, the goal of this study is to formulate
a multivariate Hawkes model to describe the news data flow as well as to
implement this framework into software. This implementation uses numerical
methods to estimate the models’ parameters given the input data. The
performance of the trained models is assessed and compared using statistical
evaluation methods, such as the likelihood and BIC measure. In addition,
this study focuses especially on testing different bucket sizes used to dis-
cretize the time scale of news arrival times and compares the results from
the different settings, e.g. by identifying the connections between different
news classes, as provided by the Hawkes model.

For the scope of this project, the underlying quality of data, e.g. how well the
aggregated news data actually represents the original news articles, will not
be analyzed. Moreover, due to limitations in computational power, some
resolution in the original data has to be removed to simplify calculations.
Likewise, in order for the model training process to converge in reasonable
time, there is some limitations in the size of the input datasets. Finally, in
CHAPTER 1. INTRODUCTION

formulating the multivariate Hawkes model, the news data is divided into classes. Though this can be done in arbitrarily many ways and levels of granularity, this study focuses on one particular composition that is used throughout the analysis.

1.3 Report Outline

This part gives an overview of the disposition of this report and the main content of each chapter. To begin with, Chapter 2 deals with news data and presents some information about aggregated news data. This part also explains the structure of the specific dataset used throughout this study. Chapter 3 presents the relevant mathematical models and algorithms. This includes some basic theory about stochastic processes and an introduction to Hawkes processes, the specific models used to describe the news data flows as well as optimization algorithms utilized in the software implementations. Next, Chapter 4 outlines the methods used in the project and focuses on the software implementation, practical handling of data and setup for the results presented in the study. Chapter 5 presents the obtained results from the analysis and Chapter 6 contains the discussion section, which ties back to the results presented in the previous part. The discussion also reflects back on the model selection, methods and implementation as well as the validity and consequences of the obtained results. Lastly, Chapter 7 presents the conclusions and summarizes the major findings of the study.
Chapter 2

News Data

A central part of this study is that of news data. This chapter first provides some general information about the characteristics of aggregated news data. Thereafter, an overview of the structure and characteristics of the specific RavenPack dataset used in this study is presented.

2.1 Aggregated News Data

For the scope of this study, the term *aggregated news data* is used to refer to news data that has been processed or altered from its original form in one way or another, typically to obtain a more compact form. For example, a text article may have been processed in a text interpretation system to extract its preamble, which can be more compactly stored in a database. This point stored in the database is then referred to as an aggregated news data point, which reflects or summarizes the content of the original article, however no longer contains all information. In addition, there exists different forms of aggregated news data. These forms depend on the original shape of the data as well as the intended use of the aggregated information. For instance, a text can be filtered using text mining techniques to extract specific fields of information, sound can be interpreted using speech recognition to be converted into text and the important events in a video can be identified by image processing methods.

There are numerous potential uses and advantages of aggregated news data. For one, the form and framework of the aggregated data can be pre-specified such that all information after filtering is given on a common form. This in turn typically makes it easier to store the data in databases, sort and filter the information based on user requirements as well as to use it for statistical analysis.
2.2 The RavenPack Data

For this study, the aggregated news data is provided by RavenPack. The dataset contains news data from January 1st 2000 until February 28th 2017. Even though the original source of each data point is an actual article or press release, i.e. news in text format, the information available in the dataset has been processed to have a more compact representation. That is, each original news piece is first processed by RavenPack and translated into their standardized framework. In this framework, each point is represented as a vector with a set of specified fields, some of which are numerical values and some categorical.

2.2.1 Dataset Overview

Before using the data in the calculations, it is important to carefully study the structure of the data as well as identify potential flaws or problems that may affect the results. To begin with, Table 2.1 below presents a list of some of the most important fields with corresponding descriptions.

\begin{table}[h]
\centering
\begin{tabular}{|l|p{0.8\textwidth}|}
\hline
\textbf{TIMESTAMP\_UTC} & A date-time string on the form YYYY-MM-DD-hh:mm:ss.sss indicating when the news data was received by the interpreting system. \\
\hline
\textbf{HEADLINE} & The headline text of the original news article. \\
\hline
\textbf{RP\_STORY\_ID} & Unique ID for each data point in the system. \\
\hline
\textbf{ENTITY\_TYPE} & The type of identified entity, which can be either Commodity, Company, Currency, Nationality, Organization, People, Place, Product or Sports teams. \\
\hline
\textbf{ENTITY\_NAME} & The name of identified entity, e.g. the name of a company or currency. \\
\hline
\textbf{COUNTRY\_CODE} & Two-character string with the ISO-3166 country code associated with the news data point, e.g. US, CH, CA. \\
\hline
\textbf{RELEVANCE} & A score taking integer values between 0 - 100 which specifies how strongly related the identified entity is to the original article, where 0 means it was passively mentioned and 100 means it was considered central to the story. \\
\hline
\end{tabular}
\caption{RavenPack dataset: important fields}
\end{table}
EVENT_SENTIMENT_SCORE
The sentiment score states how positive or negative a related event is. More specifically, it is a score between -1.00 and 1.00 with 2 decimal places that represents the news sentiment where -1.00 is very negative and 1.0 is strongly positive.

EVENT_RELEVANCE
An integer score taking values 0 - 100 that indicates the relevance of the identified event. A score of 100 means that it is important and stated in the headline, whereas a lower score means it was less central and stated further down in the article.

PROVIDER_ID
The ID of the provider of the news content, e.g. AN for Alliance News and DJ for Dow Jones Newswires.

In addition to these fields, RavenPack has a hierarchical taxonomy system to classify the content of the news data points. This particular subset of fields enables categorization and filtering on different levels of granularity. The layers in this hierarchical structure are presented in Table 2.2.

Furthermore, a key observation from the data is the existence of an event and how it affects the structure of the corresponding data point. More specifically, the data can broadly be separated into two categories; points with an identified event and points without. The points without an event contain substantially less information and lack both sentiment score as well as the hierarchical field structure given in Table 2.2. Out of the total amount of data points, 8.4% contain such an event and the related information. For this study, the points without an event are deemed to contain too little information and are therefore left out of the analysis. That is, only points containing an event and the related information are used in order to perform the desired calculations.
Table 2.2: RavenPack dataset: fields in the hierarchical taxonomy

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOPIC</td>
<td>Highest order in the classification, which can take either of the 5 labels: economy, business, society, politics or environment.</td>
</tr>
<tr>
<td>GROUP</td>
<td>Second level classifier, which has a total of 56 possible values, e.g. interest-rates, war-conflict, acquisition-mergers and earnings.</td>
</tr>
<tr>
<td>TYPE</td>
<td>Third level classifier with 495 different labels. For instance, the GROUP war-conflict has TYPE labels military-action, bombing etc.</td>
</tr>
<tr>
<td>SUB_TYPE</td>
<td>A further subdivision of the TYPE attribute. For instance, the GROUP war-conflict has SUB_TYPE labels threat, exercise etc.</td>
</tr>
<tr>
<td>PROPERTY</td>
<td>An attribute of the event, such as a role or entity. For instance, the GROUP war-conflict has PROPERTY labels attacker, location etc.</td>
</tr>
<tr>
<td>CATEGORY</td>
<td>The most detailed level, combining SUB_TYPE and PROPERTY.</td>
</tr>
</tbody>
</table>

2.2.2 Visualizations of Important Fields

In this part, a number of visualizations are presented to provide a better insight into the data characteristics and the distributions of important fields. Firstly, Figure 2.1 presents a histogram with the number of news data points from the 20 countries with the largest news flows. The countries are represented by their two-figure country code, as described in Table 2.1. One important realization here is the large over-representation of news with country code US. In addition, a total of 253 distinct country codes are present in the data set, out of which some have a very small appearance frequency.
CHAPTER 2. NEWS DATA

Figure 2.1: Number of observations for the 20 country codes with largest data flows.

Next, Figures 2.2 and 2.3 show pie charts for the distributions of \textit{RELEVANCE} and \textit{EVENT\_RELEVANCE} fields respectively. As stated in Table 2.1, all values are here given as integers.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig22.png}
\caption{Relevance \textit{R}}
\end{figure}

\begin{itemize}
\item \textbf{Red:} $0 \leq R \leq 79$,
\item \textbf{Cyan:} $80 \leq R \leq 89$,
\item \textbf{Green:} $90 \leq R \leq 99$,
\item \textbf{Blue:} $R = 100$.
\end{itemize}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig23.png}
\caption{Event Relevance \textit{ER}}
\end{figure}

\begin{itemize}
\item \textbf{Red:} $0 \leq ER \leq 79$,
\item \textbf{Cyan:} $80 \leq ER \leq 89$,
\item \textbf{Green:} $90 \leq ER \leq 99$,
\item \textbf{Blue:} $ER = 100$.
\end{itemize}

The size of the news data flow over time is examined in Figure 2.4, which presents a histogram with the yearly amount of news. Here, it is noticed that
the number of news data points increases over the years. One explanation for this is the increased availability of online news and thus, an increase in providers to the RavenPack system input. Here, it is also noted that the year 2017 only contains data points from January and February.

![Graph showing number of observations per year for years 2000 - 2017.](image)

*Figure 2.4: Number of observations per year for years 2000 - 2017.*

Other properties can be seen when examining smaller time scopes. Figure 2.5 shows the daily amount of news for the month of December in 2014. Here, a seasonality phenomenon can be seen, where the news data flow is substantially smaller during weekends in comparison to that during the week days. Additionally, the flow of news is seen to decrease over the Christmas holidays.
Next, Figure 2.6 shows the count of news data points for a selection of labels in the GROUP field. It can be seen that there are quite noticeable difference in the amounts, where labels like stock-prices have substantially higher number of observations than for instance pollution.

Lastly, Figure 2.7 shows the distribution of the EVENT_SENTIMENT_SCORE field. Note that this field only takes values between $-1.00$ and $1.00$ with a granularity of two decimal places, as described in Table 2.1. Here, 26% takes the value 0.00. In addition to this, there are two clusters of points roughly around sentiments $-0.50$ and $0.50$ respectively.
CHAPTER 2. NEWS DATA

Figure 2.6: Number of observations for 20 GROUP labels.

Figure 2.7: EVENT_SENTIMENT_SCORE distribution with interval range 0.05.
Chapter 3

Mathematical Background

In this chapter, a formal mathematical background to the models utilized in this study is given. To begin with, some theory about stochastic processes is presented. In particular, this part focuses on presenting the Hawkes process, its important properties as well as how it is different from more simple models. Thereafter, the models for the news data flows are formalized, which ties back to both the data structure presented in Chapter 2 as well as the stochastic process theory provided in the first section of this chapter. Here, the distinct classes model is the one most central to the scope of this thesis. However, a second model with overlapping classes is also introduced. Though this model is not tested in this study, it provides a generalization that could be useful for future works. The chapter also provides some background theory on the optimization algorithms and parameter estimation procedures used in the implementation. Lastly, some theory on evaluation of statistical models and model selection is provided.

3.1 Stochastic Processes

This part provides some important mathematical background on stochastic processes and the Hawkes process in particular, which is the essential part of modeling the news data flow in this thesis project. However, prior to defining the Hawkes process model, some more basic concepts are outlined.

3.1.1 Basic Stochastic Processes

Firstly, the topic of point processes is an important concept in probability theory and is especially central in modeling spatial data. In the setting of news data flow, a point process can intuitively be thought of as the random
variables describing the news arrival times. The formal definition of a point process \cite{10, 11} is stated below.

**Definition 3.1 (Point process)**
A sequence of real-valued non-negative random variables \( T = \{T_1, T_2, \ldots \} \) on a probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) is a point process if

(i) \( \mathbb{P}(0 \leq T_1 \leq T_2 \leq \ldots) = 1 \),

(ii) The number of points in a bounded region of \([0, \infty)\) is finite almost surely, i.e. \( \mathbb{P}\left( \lim_{n \to \infty} T_n = \infty \right) = 1 \).

In many cases, a point process has a corresponding count process that describes the cumulative count of arrivals. The definition of the counting process is presented below.

**Definition 3.2 (Counting process)**
A stochastic process \( N: [0, \infty) \times \Omega \to \mathbb{N}_0 \) on a probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \), with \( N_t: \Omega \to \mathbb{N}_0 \) such that \( N_t(\omega) = N(t, \omega) \quad \forall \omega \in \Omega \), is a counting process if

(i) \( \mathbb{P}(N_0 = 0) = 1 \),

(ii) \( \mathbb{P}(N_t < \infty) = 1, \quad \forall t \in [0, \infty) \),

(iii) it holds that \( N \) is a non-decreasing and right-continuous step function with increment size 1.

Furthermore, a useful concept related to the point- and counting processes is the history sigma algebra. That is, for each time \( t \in [0, \infty) \), the history sigma algebra \( \mathcal{H}_t \) of a counting process \( N \) is given as \( \mathcal{H}_t = \sigma(\{N_u: 0 \leq u \leq t\}) \).

Consequently, the sequence \( \mathcal{H} = \{\mathcal{H}_t\}_{t \in [0, \infty)} \) is a filtration on the measurable space \( (\Omega, \mathcal{F}) \). How a counting process depends on its related filtration is of great significance in many applications. Its importance for this study will be presented later. An important counting process with some special properties is the Poisson process \cite{12}. In the parts below, the definitions of both its homogeneous and inhomogeneous forms are given.

**Definition 3.3 (Homogeneous Poisson process)**
A counting process \( N: [0, \infty) \times \Omega \to \mathbb{N}_0 \) on a probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) is a homogeneous Poisson process with intensity \( \lambda \geq 0 \) if for arbitrary \( t \in [0, \infty) \) it holds for all \( h \geq 0 \) that

\[
\mathbb{P}(N_{t+h} - N_t = m) = \begin{cases} 
1 - \lambda h + \mathcal{O}(h), & m = 0, \\
\lambda h + \mathcal{O}(h), & m = 1, \\
\mathcal{O}(h), & m > 1,
\end{cases}
\]

(3.1)

where \( \mathcal{O} \) signifies some function \( o: [0, \infty) \to \mathbb{R} \) with the property
\[
\lim_{h \to 0} \frac{o(h)}{h} = 0, \tag{3.2}
\]

which also implies that \( o(0) = 0 \). This definition in turn gives that non-overlapping intervals of \( N \) are independent random variables, i.e. for all \( t \in [0, \infty) \) it holds that the increment \( N_{t+h} - N_t \) is independent of \( \mathcal{H}_t \). Furthermore, all increments are stationary and have the property such that \( N_{t+h} - N_t \sim \text{Po}(\lambda h) \). The term **homogeneous** specifies that there is no time dependency in the intensity, however in some situations it may happen that the intensity is not a constant but instead varies with time, e.g. with some linear increase or seasonal oscillations. In such a case, an **inhomogeneous** Poisson process is obtained. The definition of such a process is presented below.

**Definition 3.4 (Inhomogeneous Poisson process)**

A counting process \( N: [0, \infty) \times \Omega \to \mathbb{N}_0 \) on a probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) is an inhomogeneous Poisson process with intensity function \( \lambda: [0, \infty) \to [0, \infty) \) if for arbitrary \( t \in [0, \infty) \) it holds for all \( h \geq 0 \) that

\[
\mathbb{P}(N_{t+h} - N_t = m) = \begin{cases} 
1 - \lambda(t)h + \mathcal{O}(h), & m = 0, \\
\lambda(t)h + \mathcal{O}(h), & m = 1, \\
\mathcal{O}(h), & m > 1,
\end{cases} \tag{3.3}
\]

where as in the homogeneous case, \( \mathcal{O} \) signifies some function \( o: [0, \infty) \to \mathbb{R} \) satisfying the property in Equation 3.2. For this case, it is given that

\[
N_{t+h} - N_t \sim \text{Po} \left( \int_t^{t+h} \lambda(u) \, du \right), \quad t \in [0, \infty). \tag{3.4}
\]

### 3.1.2 The Hawkes Process

Now, it is time to formally introduce the Hawkes process. The Hawkes process is in some ways a generalization of the Poisson process, however where the process is self-exciting. This means that every observed arrival in the process causes an increase in the value of the intensity function, thus also increasing the probability of observing more arrivals in the future. In addition, this implies that the intensity does not only vary with time, but also depends on the history sigma algebra generated by the process up until the current time point. The definition of the Hawkes process \([10]\) is presented below.
Definition 3.5 (Hawkes process)
A counting process $N: [0, \infty) \times \Omega \rightarrow \mathbb{N}_0$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with associated filtration $\mathcal{H}$ is a Hawkes process if for arbitrary $t \in [0, \infty)$ it holds that

(i) for all $h \geq 0$

$$
\mathbb{P}(N_{t+h} - N_t = m \mid \mathcal{H}_t) = \begin{cases}
1 - \lambda^*(t)h + \mathcal{O}(h), & m = 0, \\
\lambda^*(t)h + \mathcal{O}(h), & m = 1, \\
\mathcal{O}(h), & m > 1,
\end{cases}
$$

(ii) the conditional intensity function $\lambda^*$ is given as

$$
\lambda^*(t) = b + \int_0^t \nu(t-u) \, dN_u, \quad t \in [0, \infty)
$$

where $b \geq 0$ is defined as the background intensity and $\nu: [0, \infty) \rightarrow [0, \infty)$ is defined as the excitation function.

As before, $\mathcal{O}$ signifies some function $o: [0, \infty) \rightarrow \mathbb{R}$ satisfying the property in Equation 3.2. Here, the conditional intensity function is an important difference from the previous Poisson process since it depends on the history of the process and so its future values are not deterministic given the current information. In a more general context, the conditional intensity function $\lambda^*$ can be defined as

$$
\lambda^*(t) = \lim_{h \searrow 0} \frac{\mathbb{E}[N_{t+h} - N_t \mid \mathcal{H}_t]}{h}, \quad t \in [0, \infty).
$$

Furthermore, the choice of excitation function $\nu$ may vary between applications and used data. One choice that has been used in for example seismological modeling is a function, also called Omori’s law, on the form

$$
\nu(t) = \frac{k}{(c + t)^p}, \quad t \in [0, \infty),
$$

where $k, p \geq 0$ and $c > 0$ are constants. Another common option is an exponential kernel on the form

$$
\nu(t) = Ve^{-\gamma t}, \quad t \in [0, \infty),
$$
where $V, \gamma$ are some non-negative constants. It can be noted that if it holds that $\nu(t) = 0 \forall t \in [0, \infty)$, the Hawkes process becomes identical to the homogeneous Poisson process. Also, for an observed sequence of arrival times $t = \{t_1, t_2, \ldots\}$ of the process during a time interval $[t_a, t_b] \subset [0, \infty)$, the conditional intensity function presented in Equation 3.6 can be written as

$$\lambda^*(t) = b + \sum_{t_l \in t; \ t_l < t} \nu(t - t_l), \quad t \in [t_a, t_b]. \quad (3.10)$$

An illustration of a Hawkes process with excitation function of the form in Equation 3.9 is given below in Figure 3.1. For this example, a time sequence $t = \{1.0, 3.0, 3.5, 4.0, 6.0\}$ is observed during the time interval $[0, 10]$ and the Hawkes parameters are set to $b = 0.1, V = 1$ and $\gamma = 1$. The upper plot shows the counting process of cumulative arrivals and the lower plot presents the conditional intensity function over the interval.

![Figure 3.1: Example of a Hawkes process and its conditional intensity function.](image)

Consequently, the likelihood function and corresponding log-likelihood function of such a realization can be written as

$$\mathcal{L}(t) = \left(\prod_{t_l \in t} \lambda^*(t_l)\right) \exp \left(- \int_{t_a}^{t_b} \lambda^*(u) \, du\right), \quad (3.11)$$
\[ \log L(t) = \sum_{t_i \in t} \log (\lambda^*(t_i)) - \int_{t_a}^{t_b} \lambda^*(u) \, du. \quad (3.12) \]

The proof for deriving this likelihood is left out of this report, however a derivation of the expression can be found in the literature reference [10].

Next, the Hawkes process can be extended to the case where multiple counting processes are considered. In such a case, the processes can have both self- and mutually-exciting properties. Such a scenario can be modeled using the multivariate Hawkes process, which is defined below.

**Definition 3.6 (Multivariate Hawkes Process)**

Consider a collection of \( n \) counting processes \( N = \{N^{(1)}, \ldots, N^{(n)}\} \) on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) with associated filtration \( \mathcal{H} \). Then \( N \) is a multivariate Hawkes process if for each \( i \in \{1, \ldots, n\} \) it holds that

(i) for all \( h \geq 0 \)

\[
P\left( N^{(i)}_{t+h} - N^{(i)}_t = m \mid \mathcal{H}_t \right) = \begin{cases} 
1 - \lambda^*_i(t)h + \mathcal{O}(h), & m = 0, \\
\lambda^*_i(t)h + \mathcal{O}(h), & m = 1, \\
\mathcal{O}(h), & m > 1, 
\end{cases} \quad (3.13)
\]

(ii) the conditional intensity function \( \lambda^*_i \) corresponding to \( N^{(i)} \) can be written on the form

\[
\lambda^*_i(t) = b_i + \sum_{j=1}^{n} \left( \int_0^t \nu_{ij}(t-u) \, dN^{(j)}_u \right), \quad t \in [0, \infty), \quad (3.14)
\]

where \( b_i \geq 0 \) is the background intensity and \( \nu_{ij} : [0, \infty) \to [0, \infty) \) is the excitation function from \( N^{(j)} \) to \( N^{(i)} \).

As before, \( \mathcal{O} \) signifies some function \( o : [0, \infty) \to \mathbb{R} \) satisfying the property in Equation 3.2. Next, consider an observed sequence of arrival times \( t \) with \( t^i = \{t^i_1, t^i_2, \ldots\} \) corresponding to each counting process \( N^{(i)}, i \in \{1, \ldots, n\} \) during a time interval \([t_a, t_b] \subset [0, \infty)\). The conditional intensity function \( \lambda^*_i \) for each \( i \) can thus be written as

\[
\lambda^*_i(t) = b_i + \sum_{j=1}^{n} \sum_{t^j_l \in t^i : t^j_l < t} \nu_{ij}(t-t^j_l), \quad t \in [t_a, t_b], \quad (3.15)
\]
with the likelihood function and corresponding log-likelihood taking the forms

\[
\mathcal{L}(t) = \prod_{i=1}^{n} \left( \prod_{t_i^j \in t_i} \lambda_i^j(t_i^j) \exp \left( - \int_{t_a}^{t_b} \lambda_i^j(u) \, du \right) \right),
\]

(3.16)

\[
\log \mathcal{L}(t) = \sum_{i=1}^{n} \left( \sum_{t_i^j \in t_i} \log \left( \lambda_i^j(t_i^j) \right) - \int_{t_a}^{t_b} \lambda_i^j(u) \, du \right),
\]

(3.17)

i.e. the total likelihood is a product over terms similar to those presented in Equation 3.11. Additionally, the exponential excitation function introduced in Equation 3.9 can be extended to the multivariate case to model the excitation from \( N(j) \) to \( N(i) \) using the form

\[
\nu_{ij}(t) = v_{ij} e^{-\gamma_i t}, \quad t \in [0, \infty),
\]

(3.18)

where \( v_{ij}, \gamma_i \) are non-negative constants, which inserted in Equation 3.14 gives the conditional intensity function for each \( i \) to take the form

\[
\lambda_i^j(t) = b_i + \sum_{j=1}^{n} \sum_{\substack{t_i^j \in t_i \colon \ t_i^j < t}} v_{ij} e^{-\gamma_i (t - t_i^j)}, \quad t \in [t_a, t_b]
\]

(3.19)

Here, \( v_{ij} \) can be thought of as elements in an excitation amplitude matrix \( V = \{v_{ij}\}_{i,j=1}^{n} \). Similarly, the parameters \( b_i \) and \( \gamma_i \) and can be thought of as elements in vectors \( b = \{b_i\}_{i=1}^{n} \) and \( \gamma = \{\gamma_i\}_{i=1}^{n} \) respectively. This is the form of the conditional intensity function that is used in this study. Of course, alternative expressions for the excitation function can also be proposed, e.g. by stating a non-stationary model where the parameters can vary with time. For instance, by redefining the background intensity constant \( b_i : [0, \infty) \to [0, \infty) \) a case similar to the one with the inhomogeneous Poisson processes presented in Definition 3.4 is obtained. This can be thought of as an inhomogeneous Hawkes process. A practical approach to this is presented in Section 4.3.3.
3.2 Modeling News Data

This section provides a description for how the news data is modeled throughout this study. To begin with, every news data point observed during some time interval \([t_a, t_b] \subseteq [0, \infty)\) is represented by a point \(y_i\) such that

\[
y_i = (t_i, x_i) \in [t_a, t_b] \times \mathcal{X},
\]

where \(t_i\) is time stamp at which the piece of news was observed, \(x_i\) is the spatial attributes of the point and \(\mathcal{X}\) is the attribute space. For instance, if the data point is described with \(m\) real-valued numerical attributes it is obtained that \(x_i \in \mathcal{X} \subseteq \mathbb{R}^m\). In this study however, there is a mixture of both real-valued numerical attributes as well as categorical attributes associated with each data point. The attribute space can therefore be written as

\[
\mathcal{X} = \mathcal{X}^{(\text{real})} \times \mathcal{X}^{(\text{cat})},
\]

where \(\mathcal{X}^{(\text{real})}\) and \(\mathcal{X}^{(\text{cat})}\) represent the real-valued and categorical dimensions of the attribute space.

Next, the sequence of observed data is defined by \(y = \{y_1, y_2, \ldots\}\) with associated arrival times and spatial attributes defined as \(t = \{t_1, t_2, \ldots\}\) and \(x = \{x_1, x_2, \ldots\}\) respectively. This data sequence includes the whole set of observed news data points, i.e. there is no sorting process based on the content of news. However, in order to properly apply the multivariate Hawkes process model, the aggregated news data ought to be partitioned into classes where each class is characterized by containing homogeneous types of news. In this study, the model for partitioning the news flow into classes have two different versions; \textit{distinct classes} and \textit{overlapping classes}. Here, the distinct classes model is the one most central to the scope of this study whereas the overlapping classes model is seen as an extension. In short, distinct classes means that the attribute space \(\mathcal{X}\) is divided into distinct subsets, where each subset corresponds to a class, whereas overlapping classes means that each class is represented by a probability density over the whole attribute space. Illustrations of these two concepts are presented below in Figures 3.2 and 3.3. Both of these two model alternatives are outlined more in detail in the next subsections.
3.2.1 Distinct Classes

In this first model it is assumed that the attribute space $\mathcal{X}$ is separated into disjoint classes. That is, if there is a total of $n$ classes it is assumed that

$$\mathcal{X} = \bigcup_{i=1}^{n} \mathcal{X}_i, \quad \mathcal{X}_i \cap \mathcal{X}_j = \emptyset, \ i \neq j, \quad (3.22)$$

where $\mathcal{X}_i$ is the part of the attribute space corresponding to class $i$. With this assumption, the flow of news data from each class $i$ is denoted as the sequence $y^i = \{y^i_1, y^i_2, \ldots \}$ with associated time sequence $t^i = \{t^i_1, t^i_2, \ldots \}$ and attribute sequence $x^i = \{x^i_1, x^i_2, \ldots \}$, similarly as in the general model but here separated by class, i.e. $y^i = \{y_j \in y: x_j \in \mathcal{X}_i\}$.

Here, the flow of news data is modeled with a multivariate Hawkes process $\mathbf{N} = \{N^{(1)}, \ldots, N^{(n)}\}$, as presented in the previous section, such that each class $i$ is represented by a counting process $N^{(i)}$ modeling the arrival times and cumulative count of news data points in that specific class. Furthermore, by using the likelihood expression stated in Equation 3.16 as well as the generalized exponential excitation function for multivariate Hawkes processes introduced in Equation 3.18 with the parameters $b, V, \gamma$ in the conditional intensity functions, the likelihood for the arrival times $t$ of an observed news data sequence $y$ during the time interval $[t_a, t_b]$ is given by

$$p(t|b, V, \gamma) = \prod_{i=1}^{n} \left( \prod_{t^i \in t} \lambda^i_t(t^i|b, V, \gamma) \right) \exp \left( - \int_{t_a}^{t_b} \lambda^i_t(t|b, V, \gamma) \, dt \right), \quad (3.23)$$
where $\lambda_i^*(t|b, V, \gamma)$ indicates the function in Equation 3.19 with the specific parameter choice $b, V, \gamma$. Likewise, $p(t|b, V, \gamma)$ is used to denote $\mathcal{L}$ given in Equation 3.16 with the parameters $b, V, \gamma$.

Next, the spatial attributes of a news data points generated in class $i$ is determined by a probability density function $f_i$:

$$X_i \xrightarrow{} [0, 1)$$

Since there is a mixture of categorical and real-valued numerical attributes, this density can for instance be partitioned into a product of multinomial densities for the categorical variables and truncated normal distribution densities for the real-valued attributes. For each class $i$, the multinomial parameters can be denoted by $\rho_i$, which represents the point-wise probabilities in the categorical domain. Similarly, the parameters of the truncated normal distribution are given by its mean $\mu_i$ and covariance $\Sigma_i$. Here, the truncated normal distribution density $f_{TN}$ for the real-valued numerical attributes taking values in $X_{i}(\text{real})$, which represents the real dimensions of $X_i$ in the same manner as in Equation 3.21, is given by

$$f_{TN}(x|\mu_i, \Sigma_i) = \frac{f_N(x|\mu_i, \Sigma_i)}{\int_{x' \in X_{i}(\text{real})} f_N(x'|\mu_i, \Sigma_i) \, dx'}, \quad x \in X_{i}(\text{real}), \quad (3.24)$$

where $f_N$ is the density function of a normal distribution such that if it is given that $X(\text{real}) \subseteq \mathbb{R}^m$ where $m \in \mathbb{N}$, then

$$f_N(x|\mu_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^m \det(\Sigma)}} \exp \left( -\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \right), \quad x \in \mathbb{R}^m. \quad (3.25)$$

Using this density model for the spatial attributes, the likelihood for an attribute sequence $x$ corresponding to an observed news data sequence $y$ can be written as

$$p(x|\rho, \mu, \Sigma) = \prod_{i=1}^{n} \prod_{x_i \in X_i} f_i(x_i|\rho_i, \mu_i, \Sigma_i), \quad (3.26)$$

where $\mu = \{\mu_i\}_{i=1}^{n}$ and $\Sigma = \{\Sigma_i\}_{i=1}^{n}$. In the same way, it can be defined that $\rho = \{\rho_i\}_{i=1}^{n}$. Next, it is modeled that the prior distribution for the parameters $b, V, \gamma, \rho, \mu, \Sigma$ can be factorized such that
Finally, given the parameterization of the time- and space factors for the news data flow as well as the factorization of the parameters’ prior distribution, the likelihood function for the observed news data sequence \( y \) also factorizes and can be written as

\[
p(y|b, V, \gamma, \rho, \mu, \Sigma) = p(t|b, V, \gamma)p(x|\rho, \mu, \Sigma).
\] (3.28)

This gives the posterior distribution over the parameters to be

\[
p(b, V, \gamma, \rho, \mu, \Sigma|y) = \frac{p(y|b, V, \gamma, \rho, \mu, \Sigma)f_{bV\gamma\rho\mu\Sigma}(b, V, \gamma, \rho, \mu, \Sigma)}{p(y)} = \frac{p(t|b, V, \gamma)f_{bV\gamma}(b, V, \gamma)\frac{p(x|\rho, \mu, \Sigma)f_{\rho\mu\Sigma}(\rho, \mu, \Sigma)}{p(x)}}{p(t)}.
\] (3.29)

From this expression, it can be concluded that the distinct class model with the presented properties yields the time- and attribute aspects to be separated in the posterior distribution, which in turn means that the time parameters and attribute parameters can be optimized independent of each other.

### 3.2.2 Overlapping Classes

A generalization of the first model would be to no longer require the attribute space to be separated into disjoint classes. In such a case, the conditional intensity function related to the Hawkes process is redefined as a function \( \lambda^* : [0, \infty) \times \mathcal{X} \to [0, \infty) \) such that for a sequence of data \( y = \{y_1, y_2, \ldots\} \) observed during the time interval \([t_a, t_b]\) it is given that

\[
\lambda^*(t, x) = b(x) + \sum_{y_i \in y: t_i < t} v(t - t_i, x, x_i), \quad t \in [t_a, t_b], \quad x \in \mathcal{X}.
\] (3.30)

In this setting, how much the news flow at a point \( x \in \mathcal{X} \) is influenced by other observations is determined by functions \( g^i : \mathcal{X} \to [0, \infty), \quad i \in \{1, \ldots, n\} \), such that each \( g^i \) is a density function that represents a class in this new
setting, which ties back to the structure introduces in Figure 3.3. Taking the sum over these densities, a function $g: \mathcal{X} \to [0, \infty)$ is defined such that

$$g(x) = \sum_{i=1}^{n} g^i(x), \quad x \in \mathcal{X}. \quad (3.31)$$

Furthermore, it is modeled that the terms in the conditional intensity function take the forms

$$b(x) = \sum_{i=1}^{n} b_i g^i(x), \quad x \in \mathcal{X}, \quad (3.32)$$

and

$$\nu(t - t', x, x') = \sum_{i=1}^{n} \sum_{j=1}^{n} V_{ij} g^i(x) \frac{g^j(x')}{g(x')} e^{-\gamma_i(t-t')},$$

$$t \in [0, \infty), \quad t' \in [0, t], \quad x, x' \in \mathcal{X}, \quad (3.33)$$

where $b_i, V_{ij}, \gamma_i$ are non-negative constants. Each element in the first sum can be interpreted as how a point $x$ in class $i$ is affected by an observed data point $x'$, weighted by the probability $\frac{g^j(x')}{g(x')}$ that point a $x'$ is in class $j$.

As shown in Figure 3.3, a point in the attribute space can be contained in several classes with different probabilities. This probability will in a general context depend on both real and categorical variables such that for each $i$ it holds that

$$g^i(x) = f(x|\rho_i, \mu_i, \Sigma_i), \quad x \in \mathcal{X}. \quad (3.34)$$

where $\rho = \{\rho_i\}_{i=1}^{n}$ represents the multinomial parameters describing the distribution over the categorical variables and $\mu = \{\mu_i\}_{i=1}^{n}$, $\Sigma = \{\Sigma_i\}_{i=1}^{n}$ represent the parameters of the truncated normal distribution used to model the distribution over the real variables. It can be noted that this setup is the same as in the distinct classes model. However, even though the overlapping and distinct models show similarities when written this way, there are important differences. A significant difference is that the spatial and time dependent parts of the likelihood no longer are independent. The
CHAPTER 3. MATHEMATICAL BACKGROUND

likelihood $p(y|b, V, \gamma, \rho, \mu, \Sigma)$ for an observed news data sequence $y$ observed during the time interval $[t_a, t_b]$ has to be written in the full form as

$$
\prod_{y_i \in Y} \lambda(t_i, x_i|b, V, \gamma, \rho, \mu, \Sigma) \exp \left( - \int_{t_a}^{t_b} \int_{x \in X} \lambda(t, x|b, V, \gamma, \rho, \mu, \Sigma) \, dx \, dt \right),
$$

(3.35)

and the posterior distribution of the parameters is obtained to be

$$
p(b, V, \gamma, \rho, \mu, \Sigma|y) = \frac{p(y|b, V, \gamma, \rho, \mu, \Sigma)f_b V \gamma \rho \mu \Sigma(b, V, \gamma, \rho, \mu, \Sigma)}{p(y)}. \tag{3.36}
$$

3.3 Optimization & Parameter Estimation

A central part of this thesis study is to estimate the parameters in the mathematical expressions that model the news data flow. More specifically, given the observations in the provided dataset and the underlying model, the likelihood function for the observed sequences can be formulated. Having stated this function, the parameters can be estimated by maximizing the likelihood with respect the these parameters in a maximum-likelihood or maximum-a-posteriori manner. However with the complex models used throughout this study, closed-form solutions for the parameters can not be formulated. In addition, many parameters need to be estimated simultaneously and the size of the input data is generally very large. Dealing with big datasets as well as high-dimensional parameter spaces is therefore of vital importance. Hence, iterative methods are used to numerically optimize the likelihood and estimate the desired parameters. This section provides some information about these numerical methods used to estimate the parameters. The most central concept here is the ADAM algorithm, which can be seen as an extension of the Gradient Descent algorithm presented below.

3.3.1 Gradient Descent

The gradient descent method \cite{13} is one of the most basic methods in numerical optimization. Consider the problem of minimizing an objective function $F: \mathbb{R}^m \to \mathbb{R}$, $m \in \mathbb{N}$. That is, the goal is to identify an optimal solution $w^* \in \mathbb{R}^m$ such that $F(w^*) \leq F(w)$, $\forall w \in \mathbb{R}^m$. Note that this is analogous to maximizing $-F$. In general, a closed-form solution for $w^*$ can not be derived. In such a case, the gradient descent algorithm can be used to find
an estimate for \( w^* \). This algorithm requires \( F \) to be differentiable and can be described by the steps given in 1.

**Algorithm 1** Gradient Descent Algorithm

1: Define convergence criteria
2: Define learning rate \( \eta_t \)
3: Initialize \( w_0 \)
4: \( t \leftarrow 0 \)
5: while not converged do:
   6: \( t \leftarrow t + 1 \)
   7: \( w_t \leftarrow w_{t-1} - \eta_t \nabla F(w_{t-1}) \)
   return \( w_t \)

Here, the learning rate \( \eta_t \) can be defined as a function of \( t \). This rate is of importance and has to be tuned to the specific problem in question in order to produce a solution that converges to the optimal value. For a suitable choice of the learning rate, the solution is guaranteed to converge to a local minimum. However, the objective function \( F \) is in general not convex, which means that the local minimum is not necessarily the global minimum. Consequently, the obtained solution will often heavily depend on the prior guess \( w_0 \).

### 3.3.2 ADAM

In some problems it is useful to adapt the algorithm to the problem-specific geometry in order to achieve faster convergence. An example of such an algorithm is ADAM [14], which can be seen as an extension to gradient descent that uses a cumulative gradient as well as an estimate for the second moment. Note that as in the gradient descent case, the aim is to minimize a differentiable objective function \( F: \mathbb{R}^m \rightarrow \mathbb{R}, m \in \mathbb{N} \) and find a point \( w^* \) such that \( F(w^*) \leq F(w), \forall w \in \mathbb{R}^m \). The procedure can be described by the steps given in Algorithm 2 below.
Algorithm 2 ADAM Algorithm

1: Define convergence criteria
2: Define step-size $\alpha$
3: Define constants $\beta_1, \beta_2 \in [0, 1)$
4: Initialize for $w_0$
5: $m_0 \leftarrow 0$ (First moment)
6: $v_0 \leftarrow 0$ (Second moment)
7: $t \leftarrow 0$
8: while not converged do:
9: $t \leftarrow t + 1$
10: $g_t \leftarrow \nabla F(w_{t-1})$
11: $m_t \leftarrow \beta_1 m_{t-1} + (1 - \beta_1) g_t$
12: $v_t \leftarrow \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$
13: $\hat{m}_t \leftarrow \frac{m_t}{1 - \beta_1^t}$
14: $\hat{v}_t \leftarrow \frac{v_t}{1 - \beta_2^t}$
15: $w_t \leftarrow w_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}$

return $w_t$

Similar to the scenario in the gradient descent algorithm with the learning rate $\eta_t$, the parameters $\beta_1, \beta_2$ and $\alpha$ defined in the ADAM algorithm have to be tuned to fit the specific problem in question.

3.4 Statistical Model Evaluation

Evaluation of statistical models is a central topic in the field of statistics. This includes both assessment for how well a specific model fits provided data as well as comparison of hierarchical models. Typically, this becomes a trade-off between choosing a more complex model, which can adapt to the data more flexibly but may cause computational issues and overfitting, or choosing a simpler model, which may be more easily handled but provide a worse fit. This is an important part of this study. Hence, this section provides some mathematical background to the statistical evaluation tests that were utilized to compare the mathematical models.

Firstly, perhaps the most fundamental concept in this area is the likelihood function [15], which has been used earlier in this report, e.g. in Equations 3.11 and 3.16. That is, given the suggested underlying model and a set of observations, the likelihood function can be stipulated. Consider a collection of parameters $\theta$ for a suggested underlying model. Let the random variables $X_1, \ldots, X_k$ have a joint density function $p(X_1, \ldots, X_k | \theta)$ based on this model. For a given sequence of observations $X_1 = x_1, \ldots, X_k = x_k$ the
likelihood function $\mathcal{L}$ is given as

$$\mathcal{L} = \mathcal{L}(x_1, \ldots, x_k) = p(x_1, \ldots, x_k|\theta).$$ \hspace{1cm} (3.37)

This means that $\mathcal{L}$ is the likelihood of the observations given that the model is true. Using this formulation, $\mathcal{L}$ can be thought of as a function of $\theta$ and thus maximized with respect to these parameters. Note that maximizing the likelihood function $\mathcal{L}$ is analogous to maximizing $\log \mathcal{L}$ by the monotonicity of the logarithmic function, or likewise to minimize $-\log \mathcal{L}$. However, the maximized likelihood function $\hat{\mathcal{L}}$ is not necessarily the best measure of assessment and cannot always be used to compare different models. This is the case since a larger model has more flexibility and will therefore always yield a larger likelihood for the same set of data than that given by a model containing a subset of the parameters in the original model. For such a case, the likelihood function says nothing about overfitting. Hence, regularization terms can be introduced to take this into account when evaluating the statistical models.

One such measure is the Bayesian information criteria, BIC [16], which takes the number of estimated parameters into account and can be used as a method for model selection. Here, an arbitrary dataset containing $k$ observations and a model with $q$ parameters can be considered. Given the maximized likelihood function $\hat{\mathcal{L}}$ obtained from the optimization step, the BIC measure is defined as

$$BIC = -2 \log \hat{\mathcal{L}} + \log(k)q.$$ \hspace{1cm} (3.38)

Thus, if several models are tested on the same dataset, their BIC values can be compared to select the model that by the BIC measure gives the best fit to the underlying data. That is, to identify which of the models that provide the smallest BIC value. In the context of this study, the negative log-likelihood and BIC measures will be the primary statistics for assessment and selection of models.
Chapter 4

Methods

In this chapter, the methods of this thesis study are presented. This includes the bucketing procedure used in this study, the implementation of the discrete classes model and setup for the training of the Hawkes models.

4.1 Bucketing

In order to computationally go through the massive amount of data, the estimation of the parameters has to be done very efficiently. As introduced in Chapter 2, the data comes at a resolution of milliseconds, which means that there are over $10^{10}$ unique time stamps for the whole dataset spanning over almost 20 years.

The behavior of the Hawkes processes is by definition dependent on the history of the process. Because of this, the implementation in this thesis uses recursive computations. However, even though this recursive method is used, looping over previous observation for every new arrival is unavoidable. Therefore, every iteration of the loop has to be done in sequence. This makes parallelism of the whole program impossible and thus, the program quickly becomes very time consuming. For this study, the proposed solution to this is to lower the resolution of the data. More specifically, if the chosen resolution is defined as 1 day, all points observed in the same day are assigned with the same time stamp. Lowering the resolution and bucketing the data in this manner thus make it possible to speed up computations.

What happens when observations are put in the same bucket is that excitation phenomena between these points are neglected. For instance, with a bucket size of 1 day, interactions that are faster than 1 day are erased in an artificial way. This means that for applications such as high-frequency trad-
ing where interaction typically occur on smaller time scales, such a bucketing procedure would likely erase a lot of useful information.

With this bucket method, a given time interval \([0, T]\) becomes a grid with intervals of increment size \(\Delta t\). In addition, \(M\) is denoted as the total number of buckets. Thus in this setting, every observation will be on one of the grid points. Likewise, this means that every grid point, or bucket, can store several events. The input time sequence \(t\) is projected on the grid according to

\[
t_G = \text{proj}_G(t) = \{\text{proj}_G(t_1), \text{proj}_G(t_2), \ldots\}, \quad (4.1)
\]

where \(t_G\) is the projected time sequence and \(\text{proj}_G\) is the operator which projects the observed time sequence onto the discretized grid of equidistant points, i.e. \(G = \{t_{M,1}, \ldots, t_{M,M}\}\). More specifically, the projection is done such that for an observed time \(t_l \in t\) it holds that

\[
\text{proj}_G(t_l) = \sup_{k \in \{1, \ldots, M\}} \{t_{M,k} : t_{M,k} \leq t_l\}. \quad (4.2)
\]

Here, the number of observations in class \(i\) and grid point with index \(k\) is denoted by \(n_{G,i}^{i,k}\). Note here that given an original time interval \([0, T]\) it holds that \(t_{M,1} = 0\) and \(t_{M,M+1} = T\). An illustration of the projection procedure is given in Figure 4.1 with the original observation times on the upper axis projected using \(M = 5\). Here, the filled bullets indicate the bucket points and the red number next to each them indicates the number of observations from the original axis that is contained in that particular bucket.

![Discretized grid and bucketing procedure.](Figure 4.1)
CHAPTER 4. METHODS

4.2 Implementation of Distinct Classes Model

Given the bucketing procedure described in the previous section and the projected sequence of times $t_G$, the integral term in the log-likelihood of the multivariate Hawkes process stated in Equation 3.16 can be written as

$$\int_0^T \lambda_i^*(u) \, du = \sum_{k=1}^{t_{M,k+1}} \int_{t_{M,k}}^{t_{M,k+1}} \lambda_i^*(u) \, du.$$  (4.3)

Consequently, the total log-likelihood of the projected time sequence $t_G$ becomes

$$\log L(t_G) = \sum_{i=1}^n \left( \sum_{k=1}^{M} n_G^{i,k} \log (\lambda_i^*(t_{M,k})) - \sum_{k=1}^{M} \int_{t_{M,k}}^{t_{M,k+1}} \lambda_i^*(u) \, du \right).$$  (4.4)

Here, it is convenient to write the intensity as a recursive sum. For each $i$ and for all $k \in \{2, \ldots, M+1\}$ it holds that

$$\lambda_i^*(t_{M,k}) = b_i + \sum_{j=1}^n \sum_{m=1}^{k-1} V_{ij} n_G^{j,m} e^{-\gamma_j(t_{M,k} - t_{M,m})}$$

$$= b_i + \sum_{j=1}^n V_{ij} \left( n_G^{j,k-1} e^{-\gamma_j(t_{M,k} - t_{M,k-1})} + \sum_{m=1}^{k-2} n_G^{j,m} e^{-\gamma_j(t_{M,k} - t_{M,m})} \right)$$

$$= b_i + \sum_{j=1}^n \left( V_{ij} n_G^{j,k-1} e^{-\gamma_j(t_{M,k} - t_{M,k-1})} + \lambda^*_{ij}(t_{M,k-1}) e^{-\gamma_j(t_{M,k} - t_{M,k-1})} \right)$$

$$= b_i + \sum_{j=1}^n \lambda^*_{ij}(t_{M,k-1}) e^{-\gamma_j(t_{M,k} - t_{M,k-1})}.$$  (4.5)

Likewise, for the point where $k = 1$ it holds that $\lambda_i^*(t_{1,M}) = \lambda_i^*(0) = b_i$. In addition, for arbitrary $u \in (t_{M,k-1}, t_{M,k}]$, $k \in \{2, \ldots, M+1\}$ it holds that

$$\lambda_i^*(u) = b_i + \sum_{j=1}^n \lambda^*_{ij}(t_{M,k-1}) e^{-\gamma_j(u - t_{M,k-1})}.$$  (4.6)
Here, the exponent label $\pm$ indicate that it is to the right of the discontinuity. Similarly, the exponent label $b$ indicates that the base intensity has been left out. The last equality comes from the definition that

$$\lambda_{ij}^{*,b,+}(t_{M,k}) = V_{ij}n_{G}^{i,k} + \lambda_{ij}^{*,b}(t_{M,k}), \quad k \in \{1, \ldots, M\}. \quad (4.7)$$

Because of the time discretization, the excitation jumps only occur at the grid points. Hence, the integral in Equation 4.4 will simply be an exponential decay scaled with $\lambda_{ij}^{*,b,+}(t_{M,k})$. It is here obtained that

$$\sum_{i=1}^{n} \left( \sum_{k=1}^{M} n_{G}^{i,k} \log(\lambda_{i}^{*}(t_{M,k})) - \sum_{k=1}^{t_{M,k+1}} \lambda_{i}^{*}(u) \, du \right)$$

$$= - \sum_{i=1}^{n} b_{i}T + \sum_{i=1}^{n} \sum_{k=1}^{M} n_{G}^{i,k} \left( \log(\lambda_{i}^{*}(t_{M,k})) - \int_{t_{M,k}}^{t_{M,k+1}} e^{-\gamma_{j}(u-t_{M,k})} \, du \right)$$

$$\sum_{j=1}^{n} \lambda_{ij}^{*,b,+}(t_{M,k}) \int_{t_{M,k}}^{t_{M,k+1}} e^{-\gamma_{j}(u-t_{M,k})} \, du$$

$$= - \sum_{i=1}^{n} b_{i}T + \sum_{i=1}^{n} \sum_{k=1}^{M} \left( n_{G}^{i,k} \log \left( b_{i} + \sum_{j=1}^{n} \lambda_{ij}^{*,b,+}(t_{M,k}) e^{-\gamma_{j}(t_{M,k+1}-t_{M,k})} \right) \right)$$

$$- \sum_{j=1}^{n} \lambda_{ij}^{*,b,+}(t_{M,k}) \frac{1 - e^{-\gamma_{j}\Delta t}}{\gamma_{j}}. \quad (4.8)$$

By rearranging the sums the expression above, the last terms is given by

$$\sum_{i=1}^{n} \sum_{k=1}^{M} \sum_{j=1}^{n} \lambda_{ij}^{*,b,+}(t_{M,k}) \frac{1 - e^{-\gamma_{j}\Delta t}}{\gamma_{j}}. \quad (4.9)$$

This means that every time step, i.e. for each term in the middle sum, a summation over the classes $j$ has to be done. When the number of classes $n$ increases, this whole process becomes very time consuming. Hence, a faster version is to rearrange the summation in the following manner

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{M} \lambda_{ij}^{*,b,+}(t_{M,k}) \frac{1 - e^{-\gamma_{j}\Delta t}}{\gamma_{j}}. \quad (4.10)$$
CHAPTER 4. METHODS

It is recalled that the only \(i\)-dependence in \(\lambda_{ij}^{*,b,+}\) is in \(V_{ij}\). Now, taking out the factor \(V_{ij}\) and defining \(\lambda_{ij}^{*,b,V,+}\) such that

\[
\lambda_{ij}^{*,b,+}(t_{M,k}) = V_{ij} \lambda_{ij}^{*,b,V,+}(t_{M,k}), \quad k \in \{1, \ldots, M\},
\]

(4.11)

the summation sequence then becomes

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} V_{ij} \sum_{k=1}^{M} \lambda_{ij}^{*,b,V,+}(t_{M,k}) \frac{1 - e^{-\gamma \Delta t}}{\gamma_j},
\]

(4.12)

which is more well-suited for parallel computations. With this, the final algorithm can be formulated. Also, using the results from Equation 3.28 that time and space attributes are separated in the distinct classes model, the total log-likelihood for an observed news data sequence \(y\) can be obtained by adding the space part of the log-likelihood at the end. Some more methods on the spatial part of the log-likelihood is presented in Section 4.3. The final algorithm is presented in Algorithm 3 below, here using \(b = \{b_i\}_{i=1}^{n}, V = \{V_{ij}\}_{i,j=1}^{n}\) and \(\gamma = \{\gamma_i\}_{i=1}^{n}\) as well as bold symbols for other vectors, e.g. defining \(\lambda_{ij}^{*,b,V,+}(t_{M,k}) = \{\lambda_{ij}^{*,b,V,+}(t_{M,k})\}_{j=1}^{n}\). For these, operations are done element-wise, except in the cases with matrix multiplication with \(V\). This is the algorithm that is implemented in the TensorFlow framework.

Algorithm 3 Discrete Classes log-likelihood with Bucketing Approximation

1: procedure LogLike
2: \(t_G \leftarrow proj_G(t)\)
3: \(\Delta t \leftarrow \frac{T}{M}\)
4: \(I_{\Delta t} \leftarrow \frac{1 - e^{-\gamma \Delta t}}{\gamma}\)
5: \(k \leftarrow 0\)
6: \(\lambda^{*,b,V,+}(t_{M,0}) \leftarrow 0\)
7: \(\textbf{while } k < M \textbf{ do:}\)
8: \(\lambda^{*,b,V}(t_{M,k}) \leftarrow \lambda^{*,b,V,+}(t_{M,k})e^{-\gamma \Delta t}\)
9: \(l_{log}(t_{M,k}) \leftarrow n_G^{t_{M,k}+1} \log (b + V \lambda^{*,V,b}(t_{M,k}))\)
10: \(\lambda^{*,b,V,+}(t_{M,k}) \leftarrow \lambda^{*,b,V}(t_{M,k}) + n_G^{t_{M,k}+1}\)
11: \(k \leftarrow k + 1\)
12: \(\log p(t_G|b, V, \gamma) \leftarrow \sum \left(-b T + \sum_{k=1}^{M} l_{log}(t_{M,k}) - V \sum_{k=1}^{M} (\lambda^{*,b,V}(t_{M,k}) I_{\Delta t})\right)\)
13: Calculate \(\log p(x|\rho, \mu, \Sigma)\)
14: \(\log p(y|b, V, \gamma, \rho, \mu, \Sigma) \leftarrow \log p(t_G|b, V, \gamma) + \log p(x|\rho, \mu, \Sigma)\)
15: \(\textbf{return } \log p(y|b, V, \gamma, \rho, \mu, \Sigma)\).
CHAPTER 4. METHODS

Given this algorithm, the negative log-likelihood is obtained by switching signs and can thereafter be used in the ADAM optimization procedure. However, since it is required that all parameters in the Hawkes process are positive, it is in practice easier to minimize over variables $\tilde{b}_i, \tilde{V}_{ij}, \tilde{\gamma}_j$ such that for all $i, j \in \{1, \ldots, n\}$ it holds that

$$b_i = \tilde{b}_i^2, \quad V_{ij} = \tilde{V}_{ij}^2, \quad \gamma_i = \tilde{\gamma}_i^2.$$  \hfill (4.13)

These new variables can take values over the whole real domain. Note that this is analogous to a maximum-a-posteriori procedure with prior distribution 1 for all these new parameters over the whole parameter space.

4.3 Setup

4.3.1 Construction of Classes

Prior to performing the calculations using the distinct classes model, the provided dataset has to divided into classes in order to obtain the structure introduced in Section 3.2. However, the construction of these classes will of course influence the quality and interpretability of the results. It is of interest to construct these classes such that homogeneous types of news end up in the same class, thus making the content of each class and connections between classes clearer.

Firstly, news data is partitioned using the GROUP field in the RavenPack framework. This level in the hierarchical taxonomy stated in Table 2.2 contains 56 unique labels, which is deemed an appropriate level of partitioning for the scope of this study.

Secondly, it is of interest to make distinctions between negative, neutral and positive news. For instance, news with GROUP label interest-rates and with positive or negative sentiment may be very different and connect to other categories of news in separate ways. Thus, the dataset is also partitioned based on the EVENT_SENTIMENT_SCORE field. Provided the original distribution presented in Figure 2.7, this is done by defining the sentiment score intervals for negative, neutral and positive news as $[-1.00, -0.30]$, $[-0.29, 0.29]$ and $[0.30, 1.00]$ respectively. This partitioning is illustrated in Figure 4.2 below.
Given this construction using both the \texttt{EVENT\_SENTIMENT\_SCORE} and \texttt{GROUP} fields to define the news classes, the total number of classes is obtained to be $3 \cdot 56 = 168$. This construction is used throughout the rest of the study in all calculations. Here, the classes are sometimes referred to by their ID, spanning from 1 to 168. A full list of class IDs and corresponding news classes is given in Appendix A.

### 4.3.2 Spatial Distribution

For the spatial part of the likelihood, truncated normal distributions introduced in Equation 3.24 are used to describe the distribution of the sentiment score within sentiment interval. The distribution over each of the sentiment intervals is used over all \texttt{GROUP} labels. That is, for an observed sequence of sentiments, here denoted $\mathbf{x} = \{x_1, x_2, \ldots \}$ since its the only spatial attribute used for the likelihood expression, it is obtained that the spatial likelihood $p(\mathbf{x}|\mu, \sigma)$ becomes

$$
\prod_{x_l \in \mathbf{x}: x_l \in \text{neg}} f_{TN}(x_l|\mu_{\text{neg}}, \sigma_{\text{neg}}) \prod_{x_l \in \mathbf{x}: x_l \in \text{neu}} f_{TN}(x_l|\mu_{\text{neu}}, \sigma_{\text{neu}}) \prod_{x_l \in \mathbf{x}: x_l \in \text{pos}} f_{TN}(x_l|\mu_{\text{pos}}, \sigma_{\text{pos}}),
$$

(4.14)
where $\sigma = \{\sigma_{\text{neg}}, \sigma_{\text{neu}}, \sigma_{\text{pos}}\}$ and $\mu = \{\mu_{\text{neg}}, \mu_{\text{neu}}, \mu_{\text{pos}}\}$ denote the one dimensional distribution parameters. In the optimization, the scale parameters $\sigma$ are only allowed to take values between 0.01 and 1 to improve algorithm stability. Here, $\text{neg}$, $\text{neu}$ and $\text{pos}$ denote the negative, neutral and positive sentiment intervals $[-1.00, 0.30]$, $[-0.29, 0.29]$ and $[0.30, 1.00]$. Also, each truncated normal distribution is defined only on its corresponding sentiment interval.

### 4.3.3 Inhomogeneous Extension

As an extension of the original multivariate Hawkes process model, a inhomogeneous version is introduced. More specifically, some non-stationary and periodic behavior in the data can be observed in Figures 2.4 and 2.5. This data structure causes issues in the original, homogeneous, Hawkes model in the sense that the excitation may primarily model the periodicity rather than more interesting connections. To handle this, an inhomogeneous background intensity is introduced, i.e. rather than describing the background intensity with constants $b_i$, these are in this case considered to be functions $b_i : [0, \infty) \to [0, \infty)$ parametrized with respect to the time $t$. In this study, this function is for each $i \in \{1, \ldots, n\}$ assumed to take the form

\[
b_i(t) = b_{i,W} \mathbb{1}_{W\text{d}}(t) + b_{i,(\text{weekend})} \mathbb{1}_{\text{Wend}}(t) + b_{i,(t)} t + b_{i,(\text{amp})} \sin \left( wt + b_{i,(\text{angle})} \right),
\]

for all $t$ in the interval of interest $[0, T]$. Here, $b_i, b_{i,(\text{weekend})}, b_{i,(t)}, b_{i,(\text{amp})}$ and $b_{i,(\text{angle})}$ are parameters that have to be estimated from data. In addition, $w = \frac{4 \pi}{365.25 \times 24 \times 60}$ min$^{-1}$ is a fixed constant that corresponds to the quarter frequency and helps modeling the economic news with periodic reporting. An example of such a news class is the GROUP label earnings, as seen in Appendix C. Also, $\mathbb{1}_{W\text{d}}$ and $\mathbb{1}_{\text{Wend}}$ are the indicator functions for times during week days and weekends respectively, i.e.

\[
\mathbb{1}_{\text{Wend}}(t) = \begin{cases} 1, & \text{if } t \text{ during weekend}, \\ 0, & \text{otherwise} \end{cases}, \quad t \in [0, T],
\]

where weekend is referred to as the days Saturday and Sunday. Also it holds that $\mathbb{1}_{W\text{d}}(t) = 1 - \mathbb{1}_{\text{Wend}}(t)$ for all $t$. The same parametrization is implemented for an inhomogeneous Poisson process, which is used to test and compare the performance of the Hawkes process models. In terms of implementation, this inhomogeneous extension only requires changing the constant $b$ in Algorithm 3 to a function of time on the form given above.
4.4 Settings & Test Scenarios

In this part, the specific settings and scenarios used in testing the algorithm are presented. The news class partitioning stated in Appendix A is used throughout all calculations. Moreover, to use datasets with observations that correspond to the stated content with a higher certainty, all data is filtered such that both \textit{RELEVANCE} and \textit{EVENT\_RELEVANCE} must take the maximum value 100.

Furthermore, a moving average tolerance level is used to define the convergence criteria and parameters are initialized using uniform random variables of suitable order of magnitude. In addition, all cases uses the same random seed to get the same initializations. The hyper-parameters used in the \textit{ADAM} are stated in Table 4.1 below.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|}
\hline
Parameter & Value \\
\hline
$\alpha$ & 1e-3 \\
$\beta_1$ & 0.9 \\
$\beta_2$ & 0.999 \\
$\epsilon$ & 1e-8 \\
\hline
\end{tabular}
\caption{ADAM algorithm hyper-parameters.}
\end{table}

The next chapter presents the results from a variety of different tests. Here, three different bucket sizes are tested; 1 day, 1 hour and 5 minutes. That is, these are the $\Delta t$ values used to construct the time grid presented above in Section 4.1. For these, the total time frames that the datasets span over are 1 year, 1 month and 1 day respectively. For instance, in the case with a bucket size of 1 day and a dataset spanning over 1 year, a total of 365 (or 366 in the case of leap year) buckets are given. Here, the yearly datasets span over the years 2012 to 2016, the monthly datasets span over the months January to May of 2015 and the daily datasets span over the days 1st - 7th March of 2015.

Moreover, for each of these settings, two types of datasets are tested; one filtered on RavenPack’s $\textit{COUNTRY\_CODE}$ field to only contain Canadian news and one without any such filtering, thus containing data points from all over the world. These cases are referred to as Canada case and World case respectively. Also, in the setting with bucket size of 1 day, the inhomogeneous background intensity extension introduced in the previous subsection is tested. For this bucket size, some more in-depth examples and visualizations of the algorithm convergence are provided. In all cases, both the multivariate Hawkes process model and Poisson model baseline are tested. It is here noted that the a Poisson model can be trained using Algorithm 3 by fixing $V = 0$. 

37
In addition, each case presents the number of observations $k$ in each dataset, optimized negative log-likelihoods as well as the obtained BIC value. Here, the BIC values are calculated using the expression in Equation 3.38. For the homogeneous models it is given that $q_{\text{Poisson}} = 174$, i.e. including the 168 process intensities and the six spatial parameters, and $q_{\text{Hawkes}} = 28566$, i.e. the parameters $b, V, \gamma$ as well as the six spatial parameters. Similarly, in the homogeneous case it is given that $q_{\text{Poisson}} = 846$ and $q_{\text{Hawkes}} = 29238$, where another 672 parameters are added to the models. Also, each case tests the performance of the trained models on subsequent datasets and compares these results.

Lastly, the connections between classes are presented. More specifically, given the trained models from the Hawkes cases, the excitation amplitudes given in the $V$ matrices are used to illustrate how the news classes relate to each other in each trained model. These connections are then compared across the different datasets and bucket sizes. Here, a filtering procedure is used to only show the largest excitations and obtain more interpretable illustrations.
Chapter 5

Results

This chapter presents the obtained results of the study. Here, news classes are sometimes referred to by their given ID. A list of all ID numbers and their corresponding class of news is presented in Appendix A. In addition, lists with the number of observation in each dataset for all news classes can be found in Appendix B-D. The negative log-likelihood values, sometimes denoted $-\log L$, refer to the value obtained from Algorithm 3 with switched signs. All values are presented with four significant figures.

5.1 Bucket Size: Day

In this section, the results for the bucket size of 1 day are presented. That is, each likelihood maximization procedure is performed on a dataset spanning over the time interval of a specific year, out of the years 2012 to 2016. Calculations are performed using both the homogeneous and inhomogeneous model. The results from each of these are provided in the subsections below. A list of the number of observations in each dataset can be found in Appendix B.

5.1.1 Homogeneous Model

Firstly, the results for the homogeneous case are presented, i.e. where homogeneous Poisson processes and multivariate Hawkes processes with constant background intensity are used. For this setup, the results for the Canadian news case are presented first. Thereafter, the results for the World case are given.
CHAPTER 5. RESULTS

Canada Case

The results presented in this part include those where homogeneous models were tested on datasets filtered for Canadian news were used. In Table 5.1 below, the number of data points $k$ as well as the minimized negative log-likelihood and corresponding BIC values are given for each of the yearly datasets stated in the left column.

Table 5.1: Results for homogeneous models on yearly Canadian news with bucket size 1 day.

<table>
<thead>
<tr>
<th>Year</th>
<th>$k$</th>
<th>$-\log \hat{L}_{\text{Poisson}}$</th>
<th>$-\log \hat{L}_{\text{Hawkes}}$</th>
<th>$\text{BIC}_{\text{Poisson}}$</th>
<th>$\text{BIC}_{\text{Hawkes}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>374968</td>
<td>1.563e6</td>
<td>1.517e6</td>
<td>3.128e6</td>
<td>3.401e6</td>
</tr>
<tr>
<td>2013</td>
<td>379130</td>
<td>1.579e6</td>
<td>1.529e6</td>
<td>3.160e6</td>
<td>3.425e6</td>
</tr>
<tr>
<td>2014</td>
<td>390248</td>
<td>1.557e6</td>
<td>1.504e6</td>
<td>3.117e6</td>
<td>3.376e6</td>
</tr>
<tr>
<td>2015</td>
<td>377274</td>
<td>1.520e6</td>
<td>1.455e6</td>
<td>3.042e6</td>
<td>3.277e6</td>
</tr>
<tr>
<td>2016</td>
<td>357837</td>
<td>1.520e6</td>
<td>1.458e6</td>
<td>3.042e6</td>
<td>3.282e6</td>
</tr>
</tbody>
</table>

From the results presented in this table, it can be concluded that the Hawkes process model provides a better likelihood for all datasets, but that all its corresponding BIC values are larger than those given by the Poisson model. From this result, the Poisson model is therefore to be favored by the BIC measure.

Next, the performances of the trained models are tested using the subsequent years’ sets of test data. For Tables 5.2 and 5.3 below, the calculated negative log-likelihood values are presented for all such combinations using both the Poisson and Hawkes models. The years for the dataset used for training are stated in the left columns and the years corresponding to the datasets used for testing are specified in the upper row. For the values presented on the diagonals, the dataset for training and testing are identical. Therefore, these negative log-likelihood values match the ones presented in Table 5.1 above.

Table 5.2: Negative log-likelihood values using homogeneous Poisson model on yearly Canadian news with bucket size 1 day.

<table>
<thead>
<tr>
<th>Year</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>1.563e6</td>
<td>1.569e6</td>
<td>1.573e6</td>
<td>1.552e6</td>
<td>1.566e6</td>
</tr>
<tr>
<td>2013</td>
<td>1.579e6</td>
<td>1.569e6</td>
<td>1.540e6</td>
<td>1.569e6</td>
<td>1.569e6</td>
</tr>
<tr>
<td>2014</td>
<td>1.557e6</td>
<td>1.552e6</td>
<td>1.578e6</td>
<td>1.565e6</td>
<td>1.565e6</td>
</tr>
<tr>
<td>2015</td>
<td>1.520e6</td>
<td>1.520e6</td>
<td>1.520e6</td>
<td>1.520e6</td>
<td>1.520e6</td>
</tr>
<tr>
<td>2016</td>
<td>1.520e6</td>
<td>1.520e6</td>
<td>1.520e6</td>
<td>1.520e6</td>
<td>1.520e6</td>
</tr>
</tbody>
</table>
From these tables above it is seen that the values on the diagonal provides the best likelihood values. This is expected as these models are trained on the same dataset as the set used for testing. In addition, it is seen that the Poisson process yields better negative log-likelihood values in all cases where the training and test sets are distinct. For the Poisson model, it is given that for each year used for test data, the model trained on the closest previous year always gives the best fit. However, no other general structure can be identified regarding whether any specific year is better or worse in the fit for the subsequent years.

Figures 5.1 and 5.2 below show the daily flow of news during the year 2015 for the news classes with ID 1 and 51 (i.e. negative acquisition-mergers and positive earnings) respectively, here shown in blue. The intensity functions given by the model trained on the data from the same year are plotted in red. In addition, the intensity functions generated by the observed data from 2015, but using the parameters from the model trained on data from 2014 are plotted in black. This procedure is conducted for the Poisson model, seen in the left subplots, as well as for the Hawkes model, which is seen in the right subplots. All intensity curves are scaled to the unit day$^{-1}$. 

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
Year & 2012 & 2013 & 2014 & 2015 & 2016 \\
\hline
2012 & 1.517e6 & 1.833e6 & 1.671e6 & 1.672e6 & 3.281e6 \\
2013 & 1.529e6 & 1.693e6 & 1.596e6 & 1.722e6 & \\
2014 & 1.504e6 & 1.738e6 & 1.959e6 & \\
2015 & 1.455e6 & 1.969e6 & \\
2016 & & & 1.458e6 & \\
\hline
\end{tabular}
\caption{Negative log-likelihood values using homogeneous Hawkes model on yearly Canadian news with bucket size 1 day.}
\end{table}
CHAPTER 5. RESULTS

Figure 5.1: Daily flow of negative acquisition-mergers news for 2015 (blue), generated intensity function from the homogeneous 2015 model (red) and generated intensity function from the homogeneous 2014 model (black) for: (a) Poisson model, (b) Hawkes model.

For the Hawkes model it is observed that the red line fits the observed data sequence quite well whereas the black line provides a worse fit and has jumps in the excitation that seem rather arbitrary. In addition, the black intensity curve fails to properly adapt to periodic behavior in the news flow for the earnings case. Since the black curve is generated by the model trained on data from the previous year, this visualization may indicate over-training of the model, which would give an explanation for the drastic jumps in the excitation.

Figure 5.2: Daily flow of positive earnings news for 2015 (blue), generated intensity function from the homogeneous 2015 model (red) and generated intensity function from the homogeneous 2014 model (black) for: (a) Poisson model, (b) Hawkes model.
World Case

In this part, the results for the World case are presented. Similar to the previous case, Table 5.4 below presents the number of observations $k$, the given negative log-likelihood values and the BIC values for each of the datasets, whose related years are stated in the left column of the table.

<table>
<thead>
<tr>
<th>Year</th>
<th>$k$</th>
<th>$-\log L_{\text{Poisson}}$</th>
<th>$-\log L_{\text{Hawkes}}$</th>
<th>$\text{BIC}_{\text{Poisson}}$</th>
<th>$\text{BIC}_{\text{Hawkes}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>6518977</td>
<td>1.188e7</td>
<td>1.145e7</td>
<td>2.376e7</td>
<td>2.336e7</td>
</tr>
<tr>
<td>2013</td>
<td>6534078</td>
<td>1.179e7</td>
<td>1.140e7</td>
<td>2.358e7</td>
<td>2.325e7</td>
</tr>
<tr>
<td>2014</td>
<td>6699944</td>
<td>1.158e7</td>
<td>1.117e7</td>
<td>2.317e7</td>
<td>2.278e7</td>
</tr>
<tr>
<td>2015</td>
<td>7143068</td>
<td>1.217e7</td>
<td>1.165e7</td>
<td>2.435e7</td>
<td>2.374e7</td>
</tr>
<tr>
<td>2016</td>
<td>6740835</td>
<td>1.261e7</td>
<td>1.189e7</td>
<td>2.521e7</td>
<td>2.424e7</td>
</tr>
</tbody>
</table>

From this table of results it is concluded that the Hawkes process model provides better log-likelihood values as well as smaller BIC values. Thus, for this case where no filtering was done on basis of country code, the Hawkes process model is the favored one by the BIC measure.

Lastly, the trained models and estimated parameters are tested on sets of test data from subsequent time intervals. In Tables 5.5 and 5.6, the years in the left column state the year of the dataset for which each model was trained and the years in the upper row state the year of the dataset for which each model was tested against. Here, the values on the diagonals match the negative log-likelihoods presented in Table 5.4 above since the datasets for model training and testing are identical in these cases.

<table>
<thead>
<tr>
<th>Year</th>
<th>Test Year</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>1.188e7</td>
<td>1.188e7</td>
<td>1.181e7</td>
<td>1.251e7</td>
<td>1.320e7</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>1.179e7</td>
<td>1.170e7</td>
<td>1.243e7</td>
<td>1.321e7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>1.158e7</td>
<td>1.227e7</td>
<td>1.318e7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>1.217e7</td>
<td>1.302e7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>1.261e7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 5. RESULTS

Table 5.6: Negative log-likelihood values using homogeneous Hawkes model on yearly World news with bucket size 1 day.

<table>
<thead>
<tr>
<th>Year</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>1.145e7</td>
<td>1.262e7</td>
<td>1.264e7</td>
<td>1.392e7</td>
<td>1.748e7</td>
</tr>
<tr>
<td>2013</td>
<td>1.140e7</td>
<td>1.278e7</td>
<td>1.232e7</td>
<td>1.359e7</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
<td>1.117e7</td>
<td>1.287e7</td>
<td>1.570e7</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td></td>
<td>1.165e7</td>
<td>1.273e7</td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.189e7</td>
</tr>
</tbody>
</table>

Here, the diagonal values always provide the best fit to the given sets of test data, which confirms the fact that the corresponding models are trained to give the best possible fit to these datasets. However, for the scenarios where training and test data are distinct, other results can be identified. As in the Canada case presented earlier, each yearly dataset used as test data is best fitted by the model trained on the previous year in the case of the Poisson model. This is however not always the case for the Hawkes process model alternative, where it is difficult to identify any general pattern in the results. One noticeable thing though is that the model trained on the year 2012 seems to provide a significantly worse fit to the years 2015 and 2016 than on 2014, which may indicate that it has become more outdated at this point. Finally, it is given that the Hawkes process model provides better log-likelihood values than the Poisson model in 2 out of the 10 cases where test and training datasets are distinct.

5.1.2 Inhomogeneous Model

With the results for the homogeneous models given in the previous subsection, it is now time to also present the results provided by the inhomogeneous models. Here, inhomogeneous Hawkes processes with background intensity functions stated on the form presented in 4.3.3 as well as inhomogeneous Poisson processes with intensity functions defined on the same form are used. In the same manner as used for the homogeneous case, the results presented in the first part below correspond to the case where datasets were filtered to only contain Canadian news. Thereafter, the results for the case with news from all over the world are presented.

Canada Case

To begin with, the results for the Canada case are presented here. In Table 5.7 shown below, the number of observations $k$, the optimized negative log-likelihood values and BIC values are stated for each dataset and corresponding year specified in the left column. It is noted that the numbers of
observations here are the same as those in Table 5.1 since the used datasets are the same.

Table 5.7: Results for inhomogeneous models on yearly Canadian news with bucket size 1 day.

<table>
<thead>
<tr>
<th>Year</th>
<th>$k$</th>
<th>$-\log \hat{L}_{\text{Poisson}}$</th>
<th>$-\log \hat{L}_{\text{Hawkes}}$</th>
<th>$\text{BIC}_{\text{Poisson}}$</th>
<th>$\text{BIC}_{\text{Hawkes}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>374968</td>
<td>1.455e6</td>
<td>1.441e6</td>
<td>2.922e6</td>
<td>3.258e6</td>
</tr>
<tr>
<td>2013</td>
<td>379130</td>
<td>1.470e6</td>
<td>1.455e6</td>
<td>2.951e6</td>
<td>3.286e6</td>
</tr>
<tr>
<td>2014</td>
<td>390248</td>
<td>1.444e6</td>
<td>1.427e6</td>
<td>2.900e6</td>
<td>3.230e6</td>
</tr>
<tr>
<td>2015</td>
<td>377274</td>
<td>1.410e6</td>
<td>1.386e6</td>
<td>2.831e6</td>
<td>3.148e6</td>
</tr>
<tr>
<td>2016</td>
<td>357837</td>
<td>1.425e6</td>
<td>1.401e6</td>
<td>2.861e6</td>
<td>3.175e6</td>
</tr>
</tbody>
</table>

From Table 5.7 above it is given that the Poisson process model gives smaller BIC values even though the negative log-likelihood values of the Hawkes alternative are smaller. Thus, for the inhomogeneous cases when using the sets with Canadian news data, the Poisson model is to be favored by the Bayesian information criteria. Furthermore, since the results presented in Tables 5.1 and 5.7 have the same underlying datasets, the table values can be compared. From this, the results indicate that the inhomogeneous Poisson process model gives better log-likelihood values than the homogeneous Hawkes process model. In addition, the inhomogeneous Poisson process model also give the smallest BIC values and is therefore the favored model in this sense among the four tested model alternatives. Lastly, between the homogeneous and inhomogeneous Hawkes model alternatives, the inhomogeneous one provides the smallest BIC values and is thus the favored choice by the BIC measure.

Next, Tables 5.8 and 5.9 below present the cases where the trained inhomogeneous models are tested against sets of test data. That each, for each model and estimated parameters trained on the datasets whose corresponding years are given in the left column, the negative log-likelihood values of the datasets whose years are stated in the upper row are presented. For the diagonal element values, the datasets for training and testing are therefore identical.

Table 5.8: Negative log-likelihood values using inhomogeneous Poisson model on yearly Canadian news with bucket size 1 day.

<table>
<thead>
<tr>
<th>Year</th>
<th>Test Year</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>2012</td>
<td>1.455e6</td>
<td>1.484e6</td>
<td>1.463e6</td>
<td>1.441e6</td>
<td>1.514e6</td>
</tr>
<tr>
<td>2013</td>
<td>2013</td>
<td>1.470e6</td>
<td>1.459e6</td>
<td>1.435e6</td>
<td>1.498e6</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>2014</td>
<td>1.444e6</td>
<td>1.429e6</td>
<td>1.523e6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>2015</td>
<td>1.410e6</td>
<td>1.507e6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>2016</td>
<td>1.425e6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For each set of test data, the diagonal elements shown in Tables 5.8 and 5.9 corresponding to the model trained on the same set of data give the best log-likelihood values. It is also given that the Poisson model alternative provides the best log-likelihood values in the cases where training and test data are distinct. Furthermore, it is possible to compare the results from these tables with the results in Tables 5.2 and 5.3 obtained from the tests with the homogeneous model. From this, it is obtained that the inhomogeneous Poisson model provides the best likelihood in all cases where training and test data are distinct. In addition, the inhomogeneous Hawkes model sometimes provides better results than the homogeneous Poisson model, though not in all cases. The homogeneous Hawkes model proves to perform the worst out of the four model alternatives.

Moreover, Figures 5.3 and 5.4 here below shows the daily flow of news data in 2015 for the news classes with ID 1 and 51 respectively, here plotted in blue. The generated intensity functions given from the model trained on the data from year 2015 are plotted in red and the generated intensity functions given from the model trained on data from year 2014 are plotted in black, both for the Poisson model shown to the left and for the Hawkes model shown to the right. The unit of the intensity curves are here scaled to the unit day\(^{-1}\).  

Table 5.9: Negative log-likelihood values using inhomogeneous Hawkes model on yearly Canadian news with bucket size 1 day.

<table>
<thead>
<tr>
<th>Year</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>1.441e6</td>
<td>1.611e6</td>
<td>1.510e6</td>
<td>1.490e6</td>
<td>1.813e6</td>
</tr>
<tr>
<td>2013</td>
<td>1.455e6</td>
<td>1.525e6</td>
<td>1.472e6</td>
<td>1.516e6</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>1.427e6</td>
<td>1.474e6</td>
<td>1.621e6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td>1.386e6</td>
<td>1.774e6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td></td>
<td></td>
<td>1.401e6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 5. RESULTS

Figure 5.3: Daily flow of negative acquisition-mergers news for 2015 (blue), generated intensity function from the inhomogeneous 2015 model (red) and generated intensity function from the inhomogeneous 2014 model (black) for: (a) Poisson model, (b) Hawkes model.

Figure 5.4: Daily flow of positive earnings news for 2015 (blue), generated intensity function from inhomogeneous 2015 model (red) and generated intensity function from inhomogeneous 2014 model (black) for: (a) Poisson model, (b) Hawkes model.

As in the homogeneous case it can be observed that for the Hawkes model it is given that the red line fits the observed flow of news data well. However, it is seen that the black line provides a worse fit. It seems to not adapt to the periodicity of the earnings news and also has jumps in the excitation that seem out of place. As previously, this is likely due to overfitting of the model generating the black line.

World Case

This part presents the results for the inhomogeneous models when tested on datasets where there is no filtration on country code. Similar to the
earlier part, Table 5.10 below states the number of observations \( k \), optimized negative log-likelihood values and BIC values for the five different datasets. The datasets’ corresponding years are stated in the left column.

<table>
<thead>
<tr>
<th>Year</th>
<th>( k )</th>
<th>(- \log \hat{\mathcal{L}}_{\text{Poisson}})</th>
<th>(- \log \hat{\mathcal{L}}_{\text{Hawkes}})</th>
<th>( \text{BIC}_{\text{Poisson}} )</th>
<th>( \text{BIC}_{\text{Hawkes}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>6518977</td>
<td>1.033e7</td>
<td>1.029e7</td>
<td>2.067e7</td>
<td>2.103e7</td>
</tr>
<tr>
<td>2013</td>
<td>6534078</td>
<td>1.024e7</td>
<td>1.020e7</td>
<td>2.050e7</td>
<td>2.086e7</td>
</tr>
<tr>
<td>2014</td>
<td>6699944</td>
<td>1.007e7</td>
<td>0.9982e7</td>
<td>2.015e7</td>
<td>2.042e7</td>
</tr>
<tr>
<td>2015</td>
<td>7143068</td>
<td>1.053e7</td>
<td>1.045e7</td>
<td>2.107e7</td>
<td>2.137e7</td>
</tr>
<tr>
<td>2016</td>
<td>6740835</td>
<td>1.117e7</td>
<td>1.103e7</td>
<td>2.236e7</td>
<td>2.251e7</td>
</tr>
</tbody>
</table>

The result table above gives that Hawkes process model has better log-likelihood values, however its BIC values are larger than those given by the Poisson model. Thus, the Poisson model is to be favored in this case by the BIC measure values. In addition, since the used datasets behind the numbers shown in Tables 5.4 and 5.10 are equivalent, its results can be compared. As in the case with Canadian news, the inhomogeneous Poisson model provides the best performance by the BIC measure. It is also seen that the inhomogeneous Poisson model gives a smaller negative log-likelihood than the homogeneous Hawkes model. Between the two Hawkes model alternatives, the inhomogeneous model gives the best performance by the BIC measure.

Next, Tables 5.11 and 5.12 give the results for the procedure of testing trained models against sets of test data from subsequent years, i.e. the negative log-likelihood values of different datasets are calculated for a range of trained models. The years of the dataset used for training are stated in left columns and the years of the dataset used as test data are stated in the upper rows. Hence, the values on the diagonals correspond to using the same datasets for training and testing. Therefore, these values match the ones presented in Table 5.10 above.

<table>
<thead>
<tr>
<th>Year</th>
<th>Test Year</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>2012</td>
<td>1.033e7</td>
<td>1.036e7</td>
<td>1.032e7</td>
<td>1.091e7</td>
<td>1.192e7</td>
</tr>
<tr>
<td>2013</td>
<td>2013</td>
<td>1.024e7</td>
<td>1.024e7</td>
<td>1.021e7</td>
<td>1.083e7</td>
<td>1.200e7</td>
</tr>
<tr>
<td>2014</td>
<td>2014</td>
<td>1.007e7</td>
<td>1.007e7</td>
<td>1.067e7</td>
<td>1.334e7</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>2015</td>
<td>1.053e7</td>
<td>1.174e7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>2016</td>
<td>1.117e7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.11: Negative log-likelihood values using inhomogeneous Poisson model on yearly World news with bucket size 1 day.
Table 5.12: Negative log-likelihood values using inhomogeneous Hawkes model on yearly World news with bucket size 1 day.

<table>
<thead>
<tr>
<th>Year</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>1.029e7</td>
<td>1.066e7</td>
<td>1.053e7</td>
<td>1.148e7</td>
<td>1.341e7</td>
</tr>
<tr>
<td>2013</td>
<td>1.020e7</td>
<td>1.093e7</td>
<td>1.201e7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>0.9982e7</td>
<td>1.098e7</td>
<td>1.259e7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td>1.045e7</td>
<td>1.176e7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.103e7</td>
</tr>
</tbody>
</table>

For the cases where training and test data are distinct, the inhomogeneous Poisson process model provides the best likelihood values in all cases apart from the one where the set for test data is from the year 2016 and the training of the model is done on the data from 2014, in which case the inhomogeneous Hawkes process model performs better. In addition, it can be seen that the Hawkes model trained on 2012 provides the worst fit for all subsequent years. It also of possible to compare these results with those obtained with the homogeneous models and the data given in Tables 5.5 and 5.6, since the underlying datasets used in each setting are the same. Examining these cases, it is clear that the inhomogeneous Poisson alternative gives the best likelihood in all cases with test data being different from the data used for training of the model. In addition, the inhomogeneous Hawkes model outperforms the homogeneous Poisson model in all cases but one. Finally, as in the case with only Canadian news data it is obtained that the homogeneous Hawkes model provides the overall worst performance.

5.1.3 Algorithm Convergence

In addition to the results presented in the subsections above, this part presents examples of how the parameter estimates and likelihood values converged in the ADAM optimization procedure. Here, Canadian news data from 2015 is used for all examples. Hence, these scenarios correspond to the results shown in Tables 5.1 and 5.7. For the parameter convergence plots, the class with ID1 is used as an example. This is the class for negative acquisition-mergers news and has a flow as presented in Figure 5.1. In Table 5.13 below, the parameters given at convergence of the algorithm are presented. The hat symbols signifies that these are the optimized parameter estimates, the term stated in brackets is the parameters unit (where \( \sim \) is used to indicate unitless measure) and an empty slot indicates no such value is present in the corresponding model.
Table 5.13: Examples of estimated parameters

<table>
<thead>
<tr>
<th>Value [Unit]</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-\log L$ [$\sim$]</td>
<td>1.520e6</td>
</tr>
<tr>
<td>$b_1$ [min$^{-1}$]</td>
<td>114.9e-5</td>
</tr>
<tr>
<td>$\gamma_1$ [min$^{-1}$]</td>
<td>7.715e-05</td>
</tr>
<tr>
<td>$V_{11}$ [min$^{-1}$]</td>
<td>9.538e-19</td>
</tr>
<tr>
<td>$V_{21}$ [min$^{-1}$]</td>
<td>1.701e-05</td>
</tr>
<tr>
<td>$\hat{\mu}_{neg}$ [$\sim$]</td>
<td>-0.5867</td>
</tr>
<tr>
<td>$\hat{\sigma}_{neg}$ [$\sim$]</td>
<td>0.1173</td>
</tr>
<tr>
<td>$b_{(weekend)}$ [min$^{-1}$]</td>
<td>3.067e-20</td>
</tr>
<tr>
<td>$b_{(t)}$ [min$^{-2}$]</td>
<td>0</td>
</tr>
<tr>
<td>$b_{(amp)}$ [min$^{-1}$]</td>
<td>5.405e-4</td>
</tr>
<tr>
<td>$b_{(angle)}$ [$\sim$]</td>
<td>3.127</td>
</tr>
</tbody>
</table>

Next, Figure 5.5 illustrates the convergences of the negative log-likelihood values for 50000 iterations of the ADAM algorithm. It can be seen that the curves converge to the values stated in Table 5.13 above. It can also be seen that the inhomogeneous models show more instability in the loss function convergence. Furthermore, Figure 5.6 shows the convergence of the stated parameters in the case of the homogeneous Hawkes process model. Once again, the plots show the updates in the parameters over 50000 iterations from initialization.

Lastly, Figure 5.7 shows the convergence of the parameters in the case of the inhomogeneous Hawkes process model using a similar setup as for the homogeneous case. Here, differences in convergence and instability between the two models can be examined. As in the case with the convergence of the negative log-likelihood, the convergence of the parameters in the inhomogeneous model is more instable. Also, it can be seen that the parameters $\mu_{neg}$ and $\sigma_{neg}$ converge in a similar manner as in the homogeneous case. This is also reflected in that the given optimized parameters presented in Table 5.13 are the same in all model cases. An example of the obtained fit to the news data is presented in Figure 5.8, showing both the empirical distributions given from the observed data, as well as the three truncated normal distributions, uniquely described by the $\mu$ and $\sigma$ parameters, corresponding to each sentiment interval.
CHAPTER 5. RESULTS

Figure 5.5: Negative log-likelihood convergence in ADAM optimization for the four model alternatives using Canadian news data with bucket size 1 day: (a) Hom.Pois, (b) Hom.Hawk, (c) Inhom.Pois, (d) Inhom.Hawk.

Figure 5.6: Parameter convergence in ADAM optimization for homogeneous Hawkes model on 2015 Canadian news data with bucket size 1 day; (a) $b_1$, (b) $\gamma_1$, (c) $V_{11}$, (d) $V_{21}$, (e) $\mu_{neg}$, (f) $\sigma_{neg}$.
Figure 5.7: Parameter convergence in ADAM optimization for inhomogeneous Hawkes model on 2015 Canadian news data with bucket size 1 day; (a) $b_1$, (b) $\gamma_1$, (c) $V_{11}$, (d) $V_{21}$, (e) $\mu_{neg}$, (f) $\sigma_{neg}$, (g) $b_1^{(weekend)}$, (h) $b_1^{(t)}$, (i) $b_1^{(amp)}$, (j) $b_1^{(angle)}$. 

52
CHAPTER 5. RESULTS

Figure 5.8: Empirical distributions of negative, neutral and positive news together with estimated truncated normal distributions using Canadian news data from 2015.

5.2 Bucket Size: Hour

This section provides the results for the cases where a time bucket size of 1 hour is used. Each dataset spans over a time interval of 1 month, which depending on the number of days in the specific month yields 672 to 744 time axis buckets. A total number of 5 datasets are used, where each dataset corresponds to a specific month between January 2015 and May 2015. A list of the number of observations in each dataset can be found in Appendix C.

The calculations are performed on two types of datasets. The first type is filtered using the `COUNTRY_CODE` label to only contain Canadian news data points. The second type has no such filtering and thus includes news data from all countries. Each type case is presented separately in the parts below.

Canada Case

Firstly, the results for the cases with datasets containing only Canadian news are presented. Each row in Table 5.14 corresponds to a dataset with its corresponding month stated in the left column. Furthermore, the number of observations $k$, the minimized negative log-likelihood values as well as the BIC values for both the Poisson and Hawkes models are presented for each separate dataset.
CHAPTER 5. RESULTS

Table 5.14: Results for homogeneous models on monthly Canadian news with bucket size 1 hour.

<table>
<thead>
<tr>
<th>Month</th>
<th>k</th>
<th>log $\hat{\mathcal{L}}_{\text{Poisson}}$</th>
<th>log $\hat{\mathcal{L}}_{\text{Hawkes}}$</th>
<th>BIC Poisson</th>
<th>BIC Hawkes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>25789</td>
<td>1.177e5</td>
<td>0.963e5</td>
<td>2.371e5</td>
<td>4.828e5</td>
</tr>
<tr>
<td>Feb</td>
<td>38147</td>
<td>1.352e5</td>
<td>1.100e5</td>
<td>2.723e5</td>
<td>5.213e5</td>
</tr>
<tr>
<td>Mar</td>
<td>37468</td>
<td>1.409e5</td>
<td>1.167e5</td>
<td>2.837e5</td>
<td>5.343e5</td>
</tr>
<tr>
<td>Apr</td>
<td>32974</td>
<td>1.287e5</td>
<td>1.053e5</td>
<td>2.593e5</td>
<td>5.078e5</td>
</tr>
<tr>
<td>May</td>
<td>41568</td>
<td>1.387e5</td>
<td>1.114e5</td>
<td>2.793e5</td>
<td>5.266e5</td>
</tr>
</tbody>
</table>

By inspecting the values presented in Table 5.14 it is seen that the Hawkes process gives better log-likelihood values, but that its larger number of parameters causes the BIC-values to be larger than those obtained from the Poisson model. Thus, in this setting the Poisson model is preferred over the Hawkes model using the Bayesian information criteria.

Next, the fit of the trained models when applied to sets of test data is evaluated. More specifically, the negative log-likelihood values are calculated using the trained models on test data from subsequent months. Each row in Tables 5.15 and 5.16 below corresponds to the same trained model obtained using the dataset from the month stated in left column. Moreover, each separate column corresponds to a set of test data, with the related month stated in the upper row. Thus, for the elements on the diagonal, the sets for training and test data are identical. It should therefore be noted that these values are the same as the negative log-likelihood values presented in Table 5.14 above.

Table 5.15: Negative log-likelihood values using homogeneous Poisson model on monthly Canadian news with bucket size 1 hour.

<table>
<thead>
<tr>
<th>Test Month</th>
<th>Month</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>1.177e5</td>
<td>1.679e5</td>
<td>1.590e5</td>
<td>1.427e5</td>
<td>1.634e5</td>
<td>1.684e5</td>
</tr>
<tr>
<td>Feb</td>
<td>1.352e5</td>
<td>1.454e5</td>
<td>1.379e5</td>
<td>1.418e5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar</td>
<td>1.409e5</td>
<td>1.349e5</td>
<td>1.441e5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr</td>
<td>1.287e5</td>
<td>1.450e5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May</td>
<td></td>
<td>1.387e5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.16: Negative log-likelihood values using homogeneous Hawkes model on monthly Canadian news with bucket size 1 hour.

<table>
<thead>
<tr>
<th>Test Month</th>
<th>Month</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>0.9632e5</td>
<td>1.718e5</td>
<td>1.898e5</td>
<td>1.682e5</td>
<td>1.748e5</td>
<td></td>
</tr>
<tr>
<td>Feb</td>
<td>1.100e5</td>
<td>1.505e5</td>
<td>1.691e5</td>
<td>1.754e5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar</td>
<td>1.167e5</td>
<td>1.716e5</td>
<td>1.890e5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr</td>
<td>1.053e5</td>
<td>1.423e5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May</td>
<td></td>
<td>1.114e5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The diagonal elements always provide the best fit since its corresponding model is already trained on the test data. In the Poisson case, the model trained on the month of January provides the worst fit to all subsequent datasets. For the Hawkes case, the model trained on the data from March also provides a slightly worse fit to the test data and the model trained on April yields a significantly better likelihood value than the other models for the subsequent set of test data. To conclude, in all cases apart from the one with a model trained on data from April and tested on the dataset May, the Poisson model provides better likelihood values when training and test sets distinct.

Below, Figures 5.9 and 5.10 present the hourly flow of negative news in classes with ID 1 and 51 plotted in blue. These also present the generated intensity functions from the model trained on the March 2015 data are shown in red and the intensity functions given from the model trained on February 2015 are shown in black. These intensity curves are scaled the unit hour$^{-1}$. The procedure is done for the both the Poisson case and the Hawkes case.

![Figure 5.9: Hourly flow of negative acquisition-mergers news for March 2015 (blue), generated intensity function from the inhomogeneous March 2015 model (red) and generated intensity function from the inhomogeneous February 2015 model (black) for: (a) Poisson model, (b) Hawkes model.](image)
CHAPTER 5. RESULTS

Figure 5.10: Flow of positive earnings news for March 2015 (blue), generated intensity function from the inhomogeneous March 2015 model (red) and generated intensity function from the inhomogeneous February 2015 model (black) for: (a) Poisson model, (b) Hawkes model.

Here, it is of interest to take a closer look at the conditional intensity functions given in the Hawkes case. From these plots it is observed that the red curves gives the overall best fit. However, the black plots also provide quite good adaptation to the observed news data flow, except from a large jump around index 600 in Figure 5.10. The model generating the black curve here seems to be able to adapt to the periodicity of the earnings news class, in contrast to the case with bucket size of 1 day. This may indicate that the resolution given with the bucket size of 1 hour provides some more information that is important to the training of the Hawkes model.

World Case

In this part, the results for the World case are presented. Similar as in the Canada case earlier, Table 5.17 below presents the number of observations \( k \) in each dataset, the optimized negative log-likelihood values and BIC values for both the Poisson and Hawkes models. Each row represents a distinct set of news data, with its corresponding month is stated in the left column.

Table 5.17: Results for homogeneous models on monthly World news with bucket size 1 hour.

<table>
<thead>
<tr>
<th>Month</th>
<th>( k )</th>
<th>(- \log \hat{L}_{\text{Poisson}})</th>
<th>(- \log \hat{L}_{\text{Hawkes}})</th>
<th>( \text{BIC}_{\text{Poisson}} )</th>
<th>( \text{BIC}_{\text{Hawkes}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>575145</td>
<td>1.033e6</td>
<td>0.7806e6</td>
<td>2.069e6</td>
<td>1.940e6</td>
</tr>
<tr>
<td>Feb</td>
<td>732552</td>
<td>0.8292e6</td>
<td>0.5363e6</td>
<td>1.661e6</td>
<td>1.458e6</td>
</tr>
<tr>
<td>Mar</td>
<td>604761</td>
<td>1.046e6</td>
<td>0.8051e6</td>
<td>2.095e6</td>
<td>1.991e6</td>
</tr>
<tr>
<td>Apr</td>
<td>658002</td>
<td>0.9954e6</td>
<td>0.7415e6</td>
<td>1.993e6</td>
<td>1.866e6</td>
</tr>
<tr>
<td>May</td>
<td>685282</td>
<td>1.002e6</td>
<td>0.7326e6</td>
<td>2.006e6</td>
<td>1.849e6</td>
</tr>
</tbody>
</table>
From this table it can be concluded that the Hawkes process yields better log-likelihood values as well as smaller BIC values when comparing it to the Poisson model. Thus, for this case where no filtering on country code was performed, the Hawkes process model is preferred over the Poisson process model when using the BIC measure.

Lastly, the obtained models are tested against sets of test data from the months following the one it was trained on. In the same way as was used in the Canada case, Tables 5.18 and 5.19 present the negative log-likelihood values for the datasets corresponding to months stated in the upper rows using the parameters estimated from the datasets from the months stated in the left columns. This means that for the values presented on the diagonal, the training dataset is the same as the test dataset and thus, the negative log-likelihood values on the diagonal match the values presented above in Table 5.17.

<table>
<thead>
<tr>
<th>Test Month</th>
<th>Month</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apr</td>
<td>Jan</td>
<td>1.033e6</td>
<td>0.9831e6</td>
<td>1.078e6</td>
<td>1.041e6</td>
<td>1.080e6</td>
</tr>
<tr>
<td></td>
<td>Feb</td>
<td>0.8292e6</td>
<td>1.158e6</td>
<td>1.073e6</td>
<td>1.043e6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mar</td>
<td>1.046e6</td>
<td>1.058e6</td>
<td>1.056e6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Apr</td>
<td>0.9954e6</td>
<td>1.047e6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>May</td>
<td>1.002e6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.19: Negative log-likelihood values using homogeneous Hawkes model on monthly World news with bucket size 1 hour.

<table>
<thead>
<tr>
<th>Test Month</th>
<th>Month</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apr</td>
<td>Jan</td>
<td>0.7806e6</td>
<td>0.8384e6</td>
<td>0.9711e6</td>
<td>0.9794e6</td>
<td>1.145e6</td>
</tr>
<tr>
<td></td>
<td>Feb</td>
<td>0.5363e6</td>
<td>1.133e6</td>
<td>1.137e6</td>
<td>1.326e6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mar</td>
<td>0.8051e6</td>
<td>1.111e6</td>
<td>1.174e6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Apr</td>
<td>0.7415e6</td>
<td>1.042e6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>May</td>
<td>0.7326e6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As before, the diagonal elements always provide the best fit to the test data. From the scenarios where the set of data points for testing is distinct from the set of training data, the Hawkes process gives a better likelihood than the Poisson model in 5 out of 10 cases. Thus, here it may be that the Hawkes model has been trained to some useful behavior in the data that causes it to be the favored model in 50% of the cases, which is better than that given in most other settings.
CHAPTER 5. RESULTS

5.3 Bucket Size: Minute

In this section, the results related to the calculations using a bucket size of 5 minutes are presented. More specifically, the likelihood optimization is performed on datasets with a total time span of 1 day, which with a bucket size of 5 minutes gives 288 time axis buckets. A total number of 7 news datasets are tested. Each such dataset corresponds to a date during the first week in March 2015, i.e. from the 1st to the 7th. Here it can be noted that the 1st of March 2015 is a Sunday, as is reflected on the smaller news flows during the 1st and 7th, i.e. during the weekend, and larger flows during the other days. As in the previous sections, the optimization is done for news datasets with country code filtering for Canada as well as for datasets where no filtering on country code was done, which corresponds to news from all over the world. The full list of observations in these datasets can be found in Appendix D.

Canada Case

Here, the results for the Canadian news datasets are presented. In Table 5.20 below, each row corresponds to a specific dataset from a certain date as denoted in the left column. Each row presents the number of data points $k$, the minimized negative log-likelihood values for the Poisson baseline case and the Hawkes case respectively as well as the BIC values for both the Poisson and Hawkes cases.

<table>
<thead>
<tr>
<th>Date</th>
<th>$k$</th>
<th>$-\log \hat{\mathcal{L}}_{\text{Poisson}}$</th>
<th>$-\log \hat{\mathcal{L}}_{\text{Hawkes}}$</th>
<th>$\text{BIC}_{\text{Poisson}}$</th>
<th>$\text{BIC}_{\text{Hawkes}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24</td>
<td>0.1372e3</td>
<td>0.1033e3</td>
<td>0.8275e3</td>
<td>90.99e3</td>
</tr>
<tr>
<td>2</td>
<td>1882</td>
<td>5.728e3</td>
<td>3.135e3</td>
<td>12.77e3</td>
<td>221.7e3</td>
</tr>
<tr>
<td>3</td>
<td>1798</td>
<td>5.918e3</td>
<td>3.877e3</td>
<td>13.14e3</td>
<td>221.8e3</td>
</tr>
<tr>
<td>4</td>
<td>1988</td>
<td>5.445e3</td>
<td>3.462e3</td>
<td>12.21e3</td>
<td>223.9e3</td>
</tr>
<tr>
<td>5</td>
<td>2517</td>
<td>5.792e3</td>
<td>3.824e3</td>
<td>12.95e3</td>
<td>231.3e3</td>
</tr>
<tr>
<td>6</td>
<td>1556</td>
<td>5.193e3</td>
<td>3.605e3</td>
<td>11.66e3</td>
<td>217.2e3</td>
</tr>
<tr>
<td>7</td>
<td>87</td>
<td>0.3932e3</td>
<td>0.3061e3</td>
<td>1.564e3</td>
<td>128.2e3</td>
</tr>
</tbody>
</table>

From this table it can be seen that though the Hawkes alternative gives a smaller value for the negative log-likelihoods, its corresponding BIC values are significantly larger for all datasets. This indicates that with the use of the BIC measure, the Poisson model is to be preferred. Here it should also be noted that the number of observations is sometimes smaller than the number of estimated parameters, particularly in the Hawkes process case.

Next, the performance of the obtained parameters from each date is tested
CHAPTER 5. RESULTS

on the news data flows of the subsequent dates. That is, the negative log-likelihood values are calculated using these parameters and are presented below in Tables 5.21 and 5.22 below for the Poisson and Hawkes cases respectively. Each row corresponds to the same set of trained parameters obtained from the dataset of the date stated in left column. In addition, each column corresponds to the same set of test data from the date stated in the upper row. This means that for the elements in the diagonal, the training and test datasets are the same and thus, the same values as presented above in Table 5.20 are given in the diagonals.

Table 5.21: Negative log-likelihood values using homogeneous Poisson model on daily Canadian news with bucket size 5 minutes.

<table>
<thead>
<tr>
<th>Date</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1372e3</td>
<td>63.68e3</td>
<td>60.81e3</td>
<td>76.75e3</td>
<td>98.02e3</td>
<td>55.48e3</td>
<td>2.838e3</td>
</tr>
<tr>
<td>2</td>
<td>5.728e3</td>
<td>12.03e3</td>
<td>13.23e3</td>
<td>13.08e3</td>
<td>8.185e3</td>
<td>4.495e3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5.918e3</td>
<td>10.07e3</td>
<td>11.62e3</td>
<td>10.46e3</td>
<td>4.161e3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5.445e3</td>
<td>8.712e3</td>
<td>9.528e3</td>
<td>3.014e3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5.792e3</td>
<td>11.96e3</td>
<td>3.632e3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5.193e3</td>
<td>7.515e3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.3932e3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.22: Negative log-likelihood values using homogeneous Hawkes model on daily Canadian news with bucket size 5 minutes.

<table>
<thead>
<tr>
<th>Date</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1032e3</td>
<td>41.80e3</td>
<td>45.76e3</td>
<td>56.60e3</td>
<td>70.25e3</td>
<td>41.51e3</td>
<td>3.057e3</td>
</tr>
<tr>
<td>2</td>
<td>3.135e3</td>
<td>26.08e3</td>
<td>47.47e3</td>
<td>25.74e3</td>
<td>32.85e3</td>
<td>7.158e3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3.877e3</td>
<td>13.54e3</td>
<td>13.17e3</td>
<td>12.73e3</td>
<td>1.938e3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.462e3</td>
<td>11.93e3</td>
<td>15.24e3</td>
<td>2.985e3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.824e3</td>
<td>14.46e3</td>
<td>3.540e3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.605e3</td>
<td>2.523e3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.3061e3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From the tables above, it can be concluded that the diagonal elements, i.e. the negative log-likelihoods with training and test data being the same, always provide the best value. For the cases where the training and test datasets are different, it is obtained that the parameters trained on the dataset from the 1st of March 2015 provides the worst fit on the 2nd to 6th of March, however it does perform better on the 7th. Moreover, apart from the Poisson case with test data as Friday the 6th of March, the model trained on the dataset from the 2nd of March performs quite poorly on the subsequent dates, especially for the Hawkes process case. The Poisson process generally provides the best test data performance, except for the model trained on the 1st of March, in which case the Hawkes process gives a better
fit in 5 out of the 6 cases. The Hawkes process also provides a better performance when using the dataset from the 7th of March as test data for 4 out of the 6 trained models.

Lastly, Figures 5.12 and 5.11 show news data flow on the 4th of March 2015 for the negative acquisition-mergers and positive earnings respectively, here shown in blue. The generated intensity functions from the model trained on data from the 4th of March 2015 are shown in red and the generated intensity functions from the model trained on data from the 3rd of March 2015 are shown in black. The left subplots show the cases for the Poisson process and the right subplots show the cases for the Hawkes process. The index on x-axes is the index of the 5-minute intervals throughout the datasets and the intensity functions are scaled to the unit of $(5 \text{ min})^{-1}$.

Figure 5.11: Flow per 5 minute interval of negative acquisition-mergers news on the 4th of March 2015 (blue), generated intensity function from homogeneous 4th of March 2015 model (red) and generated intensity function from inhomogeneous 3rd of March 2015 model (black) for: (a) Poisson model, (b) Hawkes model.
CHAPTER 5. RESULTS

Figure 5.12: Flow per 5 minute interval of positive earnings news on the 4th March 2015 (blue), generated intensity function from homogeneous 4th of March 2015 model (red) and generated intensity function from inhomogeneous 3rd of March 2015 model (black) for: (a) Poisson model, (b) Hawkes model.

In the plots corresponding to the Hawkes model in Figures 5.11 and 5.12 above it is difficult to draw any general conclusions. The red curves seem to fit a little bit better to the observed data sequence, though this is not entirely clear from visual inspection.

World Case

Next, the results for the World case are presented. As previously in the Canada case, each row in Table 5.23 below corresponds to a specific dataset from a certain date as denoted in the left column. Each row presents the number of data points $k$, the minimized negative log-likelihood values for the Poisson baseline cases and the Hawkes cases as well as the BIC values for both the Poisson and Hawkes cases.

Table 5.23: Results for homogeneous models on daily World news with bucket size 5 minutes.

<table>
<thead>
<tr>
<th>Date</th>
<th>$k$</th>
<th>$-\log \hat{L}_{\text{Poisson}}$</th>
<th>$-\log \hat{L}_{\text{Hawkes}}$</th>
<th>BIC$_{\text{Poisson}}$</th>
<th>BIC$_{\text{Hawkes}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3446</td>
<td>1.003e4</td>
<td>0.6609e4</td>
<td>2.148e4</td>
<td>24.59e4</td>
</tr>
<tr>
<td>2</td>
<td>30451</td>
<td>3.191e4</td>
<td>1.656e4</td>
<td>6.562e4</td>
<td>32.80e4</td>
</tr>
<tr>
<td>3</td>
<td>31605</td>
<td>3.245e4</td>
<td>1.919e4</td>
<td>6.670e4</td>
<td>33.44e4</td>
</tr>
<tr>
<td>4</td>
<td>33025</td>
<td>2.950e4</td>
<td>1.801e4</td>
<td>6.081e4</td>
<td>33.33e4</td>
</tr>
<tr>
<td>5</td>
<td>32147</td>
<td>3.212e4</td>
<td>2.104e4</td>
<td>6.055e4</td>
<td>33.85e4</td>
</tr>
<tr>
<td>6</td>
<td>24220</td>
<td>3.495e4</td>
<td>2.534e4</td>
<td>7.166e4</td>
<td>33.91e4</td>
</tr>
<tr>
<td>7</td>
<td>4343</td>
<td>1.175e4</td>
<td>0.9397e4</td>
<td>2.450e4</td>
<td>25.81e4</td>
</tr>
</tbody>
</table>

From this table, it can be concluded that the Hawkes model provides a better log-likelihood value. Though, its larger amount of parameters causes
CHAPTER 5. RESULTS

the BIC-values for the Hawkes process to be significantly larger. Hence by the BIC value, the Poisson process is to be preferred here.

Moreover, the obtained parameters for each trained model are tested on the subsequent news data flows, using the same procedure as in the Canada case. The negative log-likelihood values are calculated using these parameters and are presented below in Tables 5.24 and 5.25 for the Poisson and Hawkes cases respectively. Each row corresponds to the same set of trained parameters obtained from the dataset of the date stated in left column. In addition, each column corresponds to the set of test data from the date stated in the upper row. As before, this means that for the elements in the diagonal, the training and test datasets are the same. Consequently, the negative log-likelihood values presented above in Table 5.23 are also given in the diagonals here.

Table 5.24: Negative log-likelihood values using homogeneous Poisson model on daily World news with bucket size 5 minutes.

<table>
<thead>
<tr>
<th>Date</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.003e4</td>
<td>22.51e4</td>
<td>28.09e4</td>
<td>26.13e4</td>
<td>23.70e4</td>
<td>19.38e4</td>
<td>3.125e4</td>
</tr>
<tr>
<td>2</td>
<td>3.191e4</td>
<td>4.601e4</td>
<td>5.304e4</td>
<td>4.422e4</td>
<td>4.877e4</td>
<td>5.040e4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3.245e4</td>
<td>10.96e4</td>
<td>9.041e4</td>
<td>6.005e4</td>
<td>6.086e4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2.950e4</td>
<td>4.522e4</td>
<td>5.200e4</td>
<td>4.422e4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.212e4</td>
<td>9.277e4</td>
<td>4.349e4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.495e4</td>
<td>3.481e4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.175e4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.25: Negative log-likelihood values using homogeneous Hawkes model on daily World news with bucket size 5 minutes.

<table>
<thead>
<tr>
<th>Date</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6609e4</td>
<td>29.21e4</td>
<td>34.44e4</td>
<td>33.42e4</td>
<td>35.12e4</td>
<td>28.09e4</td>
<td>5.290e4</td>
</tr>
<tr>
<td>2</td>
<td>1.656e4</td>
<td>32.15e4</td>
<td>82.91e4</td>
<td>39.34e4</td>
<td>89.33e4</td>
<td>22.26e4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.919e4</td>
<td>6.676e4</td>
<td>12.51e4</td>
<td>7.738e4</td>
<td>4.380e4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.801e4</td>
<td>5.824e4</td>
<td>7.271e4</td>
<td>5.760e4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2.104e4</td>
<td>5.630e4</td>
<td>4.067e4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2.534e4</td>
<td>2.848e4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.9397e4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From these results it can once again be noted that the values on the diagonal, i.e. the cases where training and test data are the same provide the best fit. In addition, the models trained on the 1st of March provides a significantly worse fit on all sets of test data, except for the one on March 7th. In the Hawkes case, the models trained using the dataset from the 2nd of March gives a worse fit on most sets of test data in comparison to those obtained by the other models. This phenomena is not obtained for the same training
date in the Poisson case. Moreover, in the Hawkes case it is always given that for each test day, the parameters trained on the closest previous date gives the best performance. In general, the Poisson models give better likelihood values on the sets of test data, with the exception of the test data from the 7th of March for which the Hawkes process performs better in half the cases.

5.4 Connections Between Classes

An important concept in this study is the connection between the classes. That is, to identify the excitation connection from one class of news to another, which is done by examining the estimated excitation matrix $V$. Using this excitation matrix, graph networks are constructed such that each node in the network represents a class of news data and each directed edge between two nodes represents the excitation from one to the other. For the results presented in this section, all examples consider the cases where the sets of data have been filtered on country code to only contain Canadian news. In each case, a filtering process is done such that only the largest excitation amplitudes are illustrated. In all networks, yellow nodes represent classes that have excitation to them and blue nodes represent classes that have excitation from them. Also, green nodes are used to represent classes that have excitation both ways.

Firstly, Figure 5.13 below shows the excitation connections in the case of the homogeneous Hawkes models when trained on yearly sets of data. The plots show the obtained graphs from the models trained on data from year 2015 and 2016 respectively. Filtering is done such that the edges have corresponding excitation values larger than 25 day$^{-1}$. It is observed that the news classes with ID 40, 50 and 51 have a lot of excitation to them. Also, in the left subplot the class with ID 4 excite a lot of other classes, however has no excitation effect in the right subplot.

Next, Figure 5.14 shows the excitation connections using the same sets of data, but in the case of the inhomogeneous Hawkes model. Filtering is done in the same manner as before such that only the largest excitations are shown in the image. It can be seen that Figures 5.13 and 5.14 are similar but that the latter has somewhat fewer edges than the former. This would indicate that the excitation amplitude values are similar, but smaller, in the inhomogeneous model.
CHAPTER 5. RESULTS

Figure 5.13: Connections for homogeneous Hawkes model on yearly Canadian data filtered for excitation amplitude values larger than 25 day\(^{-1}\) for years: (a) 2015, (b) 2016.

Figure 5.14: Connections for inhomogeneous Hawkes model on yearly Canadian data filtered for excitation amplitude values larger than 25 day\(^{-1}\) for years: (a) 2015, (b) 2016.

A similar example is given in the case for the homogeneous Hawkes models trained on monthly data and is presented below in Figure 5.13. The two plots show the obtained graphs from the models trained on data from months March and April of 2015 respectively. The filtering here is more restrictive than in the two earlier cases and only excitation values larger than 100 day\(^{-1}\) are shown. This is also tested for the daily data from 3rd and 4th of March 2015, as can be seen in Figure 5.16. Here, filtering is done with threshold value 250 day\(^{-1}\). Noticeable differences in excitation can be seen between the two subplots. It is also observed that the news class with ID 9 has excitation...
both ways.

Figure 5.15: Connections for homogeneous Hawkes model on monthly Canadian data filtered for excitation amplitude values larger than 100 day$^{-1}$ for the months: (a) March 2015, (b) April 2015.

Figure 5.16: Connections for homogeneous Hawkes model on daily Canadian data filtered for excitation values larger than 250 day$^{-1}$ for the dates: (a) 3rd of March 2015, (b) 4th of March 2015.

It is also of interest to examine the average excitation values over the different datasets. In this case, the excitation elements are filtered such that each element must have a value above some threshold value for all utilized datasets. More specifically, if the condition holds for all models in the specific scenario, the empirical mean value is calculated for that element in the excitation matrix. Otherwise, the element value is set to zero. This is first done for the case with the homogeneous Hawkes model trained on yearly
data for the years 2012 - 2016 and is presented in Figure 5.17. Here, filtering is done such that each excitation must be larger than 0.001 day$^{-1}$ for each of the five datasets. It can be examined that only three edges remain; ID13 to ID60, ID114 to ID126 and ID143 to ID44. The same procedure is conducted for the case with inhomogeneous Hawkes model trained on yearly data for the years 2012 - 2016 and is presented in Figure 5.18. Filtering is done as earlier, i.e. such that each excitation must be larger than 0.001 day$^{-1}$ for each yearly dataset. For this scenario, only the edge ID13 to ID60 remain after filtering.

Finally, the excitation average is taken for the case with homogeneous models trained on monthly news data from Jan-May 2015 and is shown in Figure 5.19. Filtering is done using the threshold excitation value 0.1 day$^{-1}$. For this setting it is given that the news class with ID 47 has excitation to several other classes. It is also noted that the class with ID 85 has self-excitation. For the case with models trained on daily news data, no connections remained after filtering even when using a threshold value as low as 1e-5 day$^{-1}$. Therefore, no excitation average plot is given for this setting.

---

**Figure 5.17:** Average connections for homogeneous Hawkes model trained on yearly Canadian data from 2012 to 2016 filtered for excitation values larger than 0.001 day$^{-1}$ for each year.

---
CHAPTER 5. RESULTS

Figure 5.18: Average connections for inhomogeneous Hawkes model trained on yearly Canadian data from 2012 to 2016 filtered for excitation values larger than 0.001 day$^{-1}$ for each year.

Figure 5.19: Average connections for homogeneous Hawkes model trained on monthly Canadian data from January to May 2015 filtered for excitation values larger than 0.1 day$^{-1}$ for each month.
Chapter 6

Discussion

This chapter presents the discussion of the analysis and obtained results. That is, the validity, consequences, flaws and alternatives of improvement are discussed. This also includes returning to the study objectives presented in Section 1.2 as well as stating suggestions for future works in the area.

6.1 Discussion of Results

In the cases with a bucket size of 1 day and yearly datasets, as presented in Section 5.1, the Hawkes process provided better performance by the BIC measure only in the case with the homogeneous model tested on World data. In addition, the inhomogeneous Poisson model alternative provided the overall best performance, both on training and test data in both the Canada and World cases. Also, in the case with Canadian data, the Poisson models tended to perform better than the Hawkes models in both fit to training and test data. However, in the case with World data, it is obtained that the inhomogeneous Hawkes model does provide better BIC values and better performance on test data for subsequent years in comparison to the homogeneous Hawkes model. From this, it can be argued that there seem to be some deterministic periodicity in the data that is best described by the inhomogeneous models. Likewise, this periodicity seem to cause the homogeneous models to over-adapt to these phenomena, such that the observed connections may not in fact be actual excitations between news classes, but rather a consequence of the way that the algorithm compensates for the periodic behavior. Here, it can also be argued that alternative forms of the parametrization in the background intensity function could be formulated, so as to better account for this phenomenon.

Next, the setting with bucket size of 1 hour and monthly datasets was pre-
Presented. Here, the results for training data fit using the BIC measure indicated that the Hawkes model performs better in the case with World data, but that the Poisson model is preferable in the case with Canadian data. Additionally, in the Canada case the Poisson model also gives the overall best performances on test data from subsequent months. In the World case however, the Hawkes model gave the best performance on test data in half the attempted cases. This may indicate that the larger dataset size of the World data is important in training the Hawkes model, however also that there may be some granularity in these datasets, which may be important to identify Hawkes process excitations for news data and that may have been lost in the setting with bucket size of 1 day, even though these datasets contained more observations. Additionally, a natural extension here would be to implement a similar inhomogeneous model as in the setting with bucket size of 1 day. This may further improve the performance and predictive usability of the models.

For the bucket size of 5 minutes, the BIC values for the Hawkes model were substantially larger than those given by the Poisson model, both in the Canada case and the World case. In testing the fit of trained models to data from subsequent dates, the Hawkes model did perform better than the Poisson model for some combinations. However, the results varied a lot and indicated on instability due to the small number of observations in the training data. Also, it could be seen that models trained on weekend days gave worse performance than models trained on week days. Likewise, the models trained on week days yielded bad performance on the weekend tests. Furthermore, for the Hawkes models it was in most cases seen that for each test day, the model trained on the closest previous day provided the best performance. This may indicate that the model has learned some characteristic in the news data flow that remains similar the next day but then changes, thus lowering the model’s performance as time goes by. However, it is difficult to draw any certain conclusion on this given the conducted tests.

A general conclusion throughout these tests seem to be that due to the Hawkes process model’s much larger amount of parameters, datasets with a large amount of observations are required for the trained models to be useful. Hence, given the limitations in computational power and the bucketing procedure, this naturally becomes a trade-off between the size of the buckets and the time span over which the dataset stretches. To conclude, further optimization of the software implementations or use of high performance computing hardware could allow for the use of larger datasets with increased data resolution or with larger time frames, which in turn may provide better results. Moreover, another solution could be to lower the amount of parameters in the model, e.g. by requiring the excitation matrix to be of a lower rank.
CHAPTER 6. DISCUSSION

For the results with the connections between news classes, large variations between datasets was seen and little similarities could be identified. This is partly seen in that the subplots in the figures comparing separate years and months tend to be quite different, with the largest excitation in one graph sometimes being absent in the other. Thus, a case of over-fitting due to lack of observations in some news classes can be identified, which is also seen in that the news classes that in these figures have large excitation effect typically have very few observations, as can be seen in the lists in Appendices B - D. The same phenomenon can be observed in that the order of magnitude of the excitations in the plots filtered and averaged over several datasets is substantially lower than the ones observed in each separate dataset.

6.2 Returning to Scope & Objectives

The first goals of the study was to formulate and implement a multivariate Hawkes approach to model the flow of news data. This is deemed to have been accomplished quite successfully as the mathematical framework takes both arrival times and news content into account and also makes it possible to relate different categories of news to each other.

In terms of methods and implementation, this was something that, even though it consumed a lot of the time throughout the project, was performed successfully and rigorously. Though some resolution in the data is lost in the bucketing procedure, this was a necessary step in speeding up computations and to be able to use datasets of suitable sizes. Furthermore, the use of the ADAM optimization algorithm and the TensorFlow framework in Python have been suitable for the scope of this thesis. However, to further optimize computations, implementation in for instance C or C++ could be useful. Another approach would be to test a transfer learning methodology, i.e. where the model is iteratively trained on larger and larger datasets and where the parameters’ initial values in each step are defined as those given from the previous iteration.

Moreover, the evaluation of the models’ fit was presented. Here, both the negative log-likelihood and BIC measure are used to take both the fit and complexity of the models into consideration. In addition, for the distinct classes cases, the models’ fit to both training data and test data was evaluated. Also, assessment of the connections between news classes in the trained models was conducted and even though these provided indications of model over-fitting, this is an important realization that can be of use for future studies.

70
6.3 Future Work

In conclusion, this study has given an introduction to the usage of Hawkes processes the field of news data analytics. This includes both the handling and filtering of data, specification of mathematical models, extensive implementation and computational methods as well as the obtained results presented in this report. However, there is also room for further analysis and testing within the field. Therefore, this section presents a selection of extensions that at the end of this thesis project were suggested as interesting topics of future works.

Testing of the Overlapping Classes Model

Even though this study provides an explanation for the mathematical background of the overlapping classes model, no implementations or results for this model are presented in this thesis report. Hence, it would be of interest to continue working with this to get a clearer understanding of the model and how it compares to the distinct classes model in terms of results and performance.

News Flow Connection to External Factors

One important, yet perhaps questionable, assumption in the Hawkes model of the news data flows in this study is that the flow is only affected by itself. That is, the intensity functions of the stochastic processes are at its most complex parameterized with respect to time and influenced by previous observations of the processes. This means that no outer factors, such as the state of the real world, has any direct impact on the news flow model, even though an increase in news about a certain topic in reality is most likely caused by a real-world event. Likewise, the news data flow is assumed to have no effect on the state of the outer world.

Though this assumption gives a simpler model, it could be of interest to extend the model and incorporate the states of outer world in order to get a more complete understanding of the nature of news data flows. This could for instance be done by modeling the state of the world as latent variables that connects to the flow of news articles, which are the observable variables.
Using the Trained Models for Prediction of Financial Assets

This study focuses exclusively on modeling the news data flow itself. However, an interesting continuation would be to investigate what information could be used to for instance predict price movements for financial assets. Such a study could for instance include using neural networks with the trained Hawkes process models and the flow of news data as inputs.
Chapter 7

Conclusions

In this thesis study, a multivariate Hawkes process approach has been used to model flows of news data. The aggregated news data was partitioned into news classes based on subject and sentiment level, such that each class contained a similar type of news observations. With the arrival times of the news being represented by a unique element in the Hawkes process, the flow of information was given a more compact representation and connections between news classes could be formulated. A time scale discretization approach was implemented to speed up computations and special attention was paid to the bucket sizes used in constructing this grid as well as the differences in results that these implied.

Tests were conducted on several different dataset setups and the performance of the trained Hawkes processes were compared with that of Poisson processes. The results indicated that the Hawkes process gives better performance only in cases with datasets containing a lot of observations, due to its large amount of unknown parameters that have to be estimated from training data. In the majority of the analyzed cases, the Poisson model provided better performance when the trained models are tested on subsequent news data flows. However, the Hawkes model showed slightly better test data performance in the setting with bucket size 1 hour, where it was the preferable model in 50% of the attempted tests. Also, it was obtained that the connections between the news classes varied substantially between datasets and that cases of overfitting occurred due to lack of observations in some of these classes. This in turn lead to models with lower performance on test data. Finally, it was suggested that a future model also ought to better account for the periodicity and deterministic behaviors in the news data flows.
Bibliography


Appendix A

List of IDs and News Classes

Table A.1: ID numbers and corresponding news classes with GROUP label and EVENT_SENTIMENT_SCORE interval.

<table>
<thead>
<tr>
<th>ID</th>
<th>GROUP label (EVENT_SENTIMENT_SCORE interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID1</td>
<td>acquisitions-mergers (negative)</td>
</tr>
<tr>
<td>ID2</td>
<td>acquisitions-mergers (neutral)</td>
</tr>
<tr>
<td>ID3</td>
<td>acquisitions-mergers (positive)</td>
</tr>
<tr>
<td>ID4</td>
<td>aid (negative)</td>
</tr>
<tr>
<td>ID5</td>
<td>aid (neutral)</td>
</tr>
<tr>
<td>ID6</td>
<td>aid (positive)</td>
</tr>
<tr>
<td>ID7</td>
<td>analyst-ratings (negative)</td>
</tr>
<tr>
<td>ID8</td>
<td>analyst-ratings (neutral)</td>
</tr>
<tr>
<td>ID9</td>
<td>analyst-ratings (positive)</td>
</tr>
<tr>
<td>ID10</td>
<td>assets (negative)</td>
</tr>
<tr>
<td>ID11</td>
<td>assets (neutral)</td>
</tr>
<tr>
<td>ID12</td>
<td>assets (positive)</td>
</tr>
<tr>
<td>ID13</td>
<td>balance-of-payments (negative)</td>
</tr>
<tr>
<td>ID14</td>
<td>balance-of-payments (neutral)</td>
</tr>
<tr>
<td>ID15</td>
<td>balance-of-payments (positive)</td>
</tr>
<tr>
<td>ID16</td>
<td>bankruptcy (negative)</td>
</tr>
<tr>
<td>ID17</td>
<td>bankruptcy (neutral)</td>
</tr>
<tr>
<td>ID18</td>
<td>bankruptcy (positive)</td>
</tr>
<tr>
<td>ID19</td>
<td>business-activity (negative)</td>
</tr>
<tr>
<td>ID20</td>
<td>business-activity (neutral)</td>
</tr>
<tr>
<td>ID21</td>
<td>business-activity (positive)</td>
</tr>
<tr>
<td>ID22</td>
<td>civil-unrest (negative)</td>
</tr>
<tr>
<td>ID23</td>
<td>civil-unrest (neutral)</td>
</tr>
<tr>
<td>ID24</td>
<td>civil-unrest (positive)</td>
</tr>
<tr>
<td>ID25</td>
<td>commodity-prices (negative)</td>
</tr>
<tr>
<td>ID26</td>
<td>commodity-prices (neutral)</td>
</tr>
<tr>
<td>ID27</td>
<td>commodity-prices (positive)</td>
</tr>
<tr>
<td>ID28</td>
<td>consumption (negative)</td>
</tr>
<tr>
<td>ID29</td>
<td>consumption (neutral)</td>
</tr>
<tr>
<td>ID30</td>
<td>consumption (positive)</td>
</tr>
<tr>
<td>ID31</td>
<td>corporate-responsibility (negative)</td>
</tr>
<tr>
<td>ID32</td>
<td>corporate-responsibility (neutral)</td>
</tr>
<tr>
<td>ID33</td>
<td>corporate-responsibility (positive)</td>
</tr>
<tr>
<td>ID34</td>
<td>credit (negative)</td>
</tr>
<tr>
<td>ID35</td>
<td>credit (neutral)</td>
</tr>
<tr>
<td>ID36</td>
<td>credit (positive)</td>
</tr>
<tr>
<td>ID37</td>
<td>credit-ratings (negative)</td>
</tr>
<tr>
<td>ID38</td>
<td>credit-ratings (neutral)</td>
</tr>
<tr>
<td>ID39</td>
<td>credit-ratings (positive)</td>
</tr>
<tr>
<td>ID40</td>
<td>crime (negative)</td>
</tr>
<tr>
<td>ID41</td>
<td>crime (neutral)</td>
</tr>
<tr>
<td>ID42</td>
<td>crime (positive)</td>
</tr>
<tr>
<td>ID43</td>
<td>dividends (negative)</td>
</tr>
<tr>
<td>ID44</td>
<td>dividends (neutral)</td>
</tr>
<tr>
<td>ID45</td>
<td>dividends (positive)</td>
</tr>
<tr>
<td>ID46</td>
<td>domestic-product (negative)</td>
</tr>
<tr>
<td>ID47</td>
<td>domestic-product (neutral)</td>
</tr>
<tr>
<td>ID48</td>
<td>domestic-product (positive)</td>
</tr>
<tr>
<td>ID</td>
<td>Term</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------</td>
</tr>
<tr>
<td>49</td>
<td>earnings</td>
</tr>
<tr>
<td>50</td>
<td>earnings</td>
</tr>
<tr>
<td>51</td>
<td>earnings</td>
</tr>
<tr>
<td>52</td>
<td>economic-union</td>
</tr>
<tr>
<td>53</td>
<td>economic-union</td>
</tr>
<tr>
<td>54</td>
<td>economic-union</td>
</tr>
<tr>
<td>55</td>
<td>elections</td>
</tr>
<tr>
<td>56</td>
<td>elections</td>
</tr>
<tr>
<td>57</td>
<td>elections</td>
</tr>
<tr>
<td>58</td>
<td>employment</td>
</tr>
<tr>
<td>59</td>
<td>employment</td>
</tr>
<tr>
<td>60</td>
<td>employment</td>
</tr>
<tr>
<td>61</td>
<td>equity-actions</td>
</tr>
<tr>
<td>62</td>
<td>equity-actions</td>
</tr>
<tr>
<td>63</td>
<td>equity-actions</td>
</tr>
<tr>
<td>64</td>
<td>exploration</td>
</tr>
<tr>
<td>65</td>
<td>exploration</td>
</tr>
<tr>
<td>66</td>
<td>exploration</td>
</tr>
<tr>
<td>67</td>
<td>foreign-exchange</td>
</tr>
<tr>
<td>68</td>
<td>foreign-exchange</td>
</tr>
<tr>
<td>69</td>
<td>foreign-exchange</td>
</tr>
<tr>
<td>70</td>
<td>foreign-relations</td>
</tr>
<tr>
<td>71</td>
<td>foreign-relations</td>
</tr>
<tr>
<td>72</td>
<td>foreign-relations</td>
</tr>
<tr>
<td>73</td>
<td>government</td>
</tr>
<tr>
<td>74</td>
<td>government</td>
</tr>
<tr>
<td>75</td>
<td>government</td>
</tr>
<tr>
<td>76</td>
<td>health</td>
</tr>
<tr>
<td>77</td>
<td>health</td>
</tr>
<tr>
<td>78</td>
<td>health</td>
</tr>
<tr>
<td>79</td>
<td>housing</td>
</tr>
<tr>
<td>80</td>
<td>housing</td>
</tr>
<tr>
<td>81</td>
<td>housing</td>
</tr>
<tr>
<td>82</td>
<td>indexes</td>
</tr>
<tr>
<td>83</td>
<td>indexes</td>
</tr>
<tr>
<td>84</td>
<td>indexes</td>
</tr>
<tr>
<td>85</td>
<td>industrial-accidents</td>
</tr>
<tr>
<td>86</td>
<td>industrial-accidents</td>
</tr>
<tr>
<td>87</td>
<td>industrial-accidents</td>
</tr>
<tr>
<td>88</td>
<td>insider-trading</td>
</tr>
<tr>
<td>89</td>
<td>insider-trading</td>
</tr>
<tr>
<td>90</td>
<td>insider-trading</td>
</tr>
<tr>
<td>91</td>
<td>interest-rates</td>
</tr>
<tr>
<td>92</td>
<td>interest-rates</td>
</tr>
<tr>
<td>93</td>
<td>interest-rates</td>
</tr>
<tr>
<td>94</td>
<td>inventory</td>
</tr>
<tr>
<td>95</td>
<td>inventory</td>
</tr>
<tr>
<td>96</td>
<td>inventory</td>
</tr>
<tr>
<td>97</td>
<td>investor-relations</td>
</tr>
<tr>
<td>98</td>
<td>investor-relations</td>
</tr>
<tr>
<td>99</td>
<td>investor-relations</td>
</tr>
<tr>
<td>100</td>
<td>labor-issues</td>
</tr>
<tr>
<td>101</td>
<td>labor-issues</td>
</tr>
<tr>
<td>102</td>
<td>labor-issues</td>
</tr>
<tr>
<td>103</td>
<td>legal</td>
</tr>
<tr>
<td>104</td>
<td>legal</td>
</tr>
<tr>
<td>105</td>
<td>legal</td>
</tr>
<tr>
<td>106</td>
<td>marketing</td>
</tr>
<tr>
<td>107</td>
<td>marketing</td>
</tr>
<tr>
<td>108</td>
<td>marketing</td>
</tr>
<tr>
<td>109</td>
<td>migration</td>
</tr>
<tr>
<td>110</td>
<td>migration</td>
</tr>
<tr>
<td>111</td>
<td>migration</td>
</tr>
<tr>
<td>112</td>
<td>natural-disasters</td>
</tr>
<tr>
<td>113</td>
<td>natural-disasters</td>
</tr>
<tr>
<td>114</td>
<td>natural-disasters</td>
</tr>
<tr>
<td>115</td>
<td>order-imbalance</td>
</tr>
<tr>
<td>116</td>
<td>order-imbalance</td>
</tr>
<tr>
<td>117</td>
<td>order-imbalance</td>
</tr>
<tr>
<td>118</td>
<td>partnerships</td>
</tr>
<tr>
<td>119</td>
<td>partnerships</td>
</tr>
<tr>
<td>120</td>
<td>partnerships</td>
</tr>
<tr>
<td>121</td>
<td>pollution</td>
</tr>
<tr>
<td>122</td>
<td>pollution</td>
</tr>
<tr>
<td>123</td>
<td>pollution</td>
</tr>
<tr>
<td>124</td>
<td>price-targets</td>
</tr>
<tr>
<td>125</td>
<td>price-targets</td>
</tr>
<tr>
<td>126</td>
<td>price-targets</td>
</tr>
<tr>
<td>127</td>
<td>production</td>
</tr>
<tr>
<td>ID</td>
<td>Category</td>
</tr>
<tr>
<td>-----</td>
<td>---------------------------</td>
</tr>
<tr>
<td>ID128</td>
<td>production (neutral)</td>
</tr>
<tr>
<td>ID129</td>
<td>production (positive)</td>
</tr>
<tr>
<td>ID130</td>
<td>products-services (negative)</td>
</tr>
<tr>
<td>ID131</td>
<td>products-services (neutral)</td>
</tr>
<tr>
<td>ID132</td>
<td>products-services (positive)</td>
</tr>
<tr>
<td>ID133</td>
<td>public-finance (negative)</td>
</tr>
<tr>
<td>ID134</td>
<td>public-finance (neutral)</td>
</tr>
<tr>
<td>ID135</td>
<td>public-finance (positive)</td>
</tr>
<tr>
<td>ID136</td>
<td>public-opinion (negative)</td>
</tr>
<tr>
<td>ID137</td>
<td>public-opinion (neutral)</td>
</tr>
<tr>
<td>ID138</td>
<td>public-opinion (positive)</td>
</tr>
<tr>
<td>ID139</td>
<td>regulatory (negative)</td>
</tr>
<tr>
<td>ID140</td>
<td>regulatory (neutral)</td>
</tr>
<tr>
<td>ID141</td>
<td>regulatory (positive)</td>
</tr>
<tr>
<td>ID142</td>
<td>revenues (negative)</td>
</tr>
<tr>
<td>ID143</td>
<td>revenues (neutral)</td>
</tr>
<tr>
<td>ID144</td>
<td>revenues (positive)</td>
</tr>
<tr>
<td>ID145</td>
<td>security (negative)</td>
</tr>
<tr>
<td>ID146</td>
<td>security (neutral)</td>
</tr>
<tr>
<td>ID147</td>
<td>security (positive)</td>
</tr>
<tr>
<td>ID148</td>
<td>social-relations (negative)</td>
</tr>
<tr>
<td>ID149</td>
<td>social-relations (neutral)</td>
</tr>
<tr>
<td>ID150</td>
<td>social-relations (positive)</td>
</tr>
<tr>
<td>ID151</td>
<td>stock-picks (negative)</td>
</tr>
<tr>
<td>ID152</td>
<td>stock-picks (neutral)</td>
</tr>
<tr>
<td>ID153</td>
<td>stock-picks (positive)</td>
</tr>
<tr>
<td>ID154</td>
<td>stock-prices (negative)</td>
</tr>
<tr>
<td>ID155</td>
<td>stock-prices (neutral)</td>
</tr>
<tr>
<td>ID156</td>
<td>stock-prices (positive)</td>
</tr>
<tr>
<td>ID157</td>
<td>taxes (negative)</td>
</tr>
<tr>
<td>ID158</td>
<td>taxes (neutral)</td>
</tr>
<tr>
<td>ID159</td>
<td>taxes (positive)</td>
</tr>
<tr>
<td>ID160</td>
<td>technical-analysis (negative)</td>
</tr>
<tr>
<td>ID161</td>
<td>technical-analysis (neutral)</td>
</tr>
<tr>
<td>ID162</td>
<td>technical-analysis (positive)</td>
</tr>
<tr>
<td>ID163</td>
<td>transportation (negative)</td>
</tr>
<tr>
<td>ID164</td>
<td>transportation (neutral)</td>
</tr>
<tr>
<td>ID165</td>
<td>transportation (positive)</td>
</tr>
<tr>
<td>ID166</td>
<td>war-conflict (negative)</td>
</tr>
<tr>
<td>ID167</td>
<td>war-conflict (neutral)</td>
</tr>
<tr>
<td>ID168</td>
<td>war-conflict (positive)</td>
</tr>
</tbody>
</table>
## Appendix B

### Observations per News Class for Yearly Datasets

<table>
<thead>
<tr>
<th>ID</th>
<th>Canada Data</th>
<th>World Data</th>
<th>Canada Data</th>
<th>World Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID1</td>
<td>699</td>
<td>459</td>
<td>730</td>
<td>684</td>
</tr>
<tr>
<td>ID2</td>
<td>2685</td>
<td>2105</td>
<td>1935</td>
<td>1551</td>
</tr>
<tr>
<td>ID3</td>
<td>20369</td>
<td>20511</td>
<td>19785</td>
<td>15264</td>
</tr>
<tr>
<td>ID4</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ID5</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>ID6</td>
<td>36</td>
<td>99</td>
<td>34</td>
<td>4</td>
</tr>
<tr>
<td>ID7</td>
<td>1361</td>
<td>2372</td>
<td>1604</td>
<td>2173</td>
</tr>
<tr>
<td>ID8</td>
<td>1019</td>
<td>1452</td>
<td>1179</td>
<td>1209</td>
</tr>
<tr>
<td>ID9</td>
<td>3370</td>
<td>2510</td>
<td>2382</td>
<td>2605</td>
</tr>
<tr>
<td>ID10</td>
<td>1385</td>
<td>2414</td>
<td>1039</td>
<td>1066</td>
</tr>
<tr>
<td>ID11</td>
<td>1385</td>
<td>2414</td>
<td>1039</td>
<td>1066</td>
</tr>
<tr>
<td>ID12</td>
<td>3046</td>
<td>2613</td>
<td>2209</td>
<td>2043</td>
</tr>
<tr>
<td>ID13</td>
<td>221</td>
<td>122</td>
<td>171</td>
<td>195</td>
</tr>
<tr>
<td>ID14</td>
<td>225</td>
<td>162</td>
<td>192</td>
<td>167</td>
</tr>
<tr>
<td>ID15</td>
<td>216</td>
<td>255</td>
<td>214</td>
<td>212</td>
</tr>
<tr>
<td>ID16</td>
<td>35</td>
<td>54</td>
<td>34</td>
<td>150</td>
</tr>
<tr>
<td>ID17</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>ID18</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>ID19</td>
<td>21</td>
<td>10</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>ID20</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ID21</td>
<td>8</td>
<td>25</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>ID22</td>
<td>497</td>
<td>551</td>
<td>249</td>
<td>292</td>
</tr>
<tr>
<td>ID23</td>
<td>31</td>
<td>103</td>
<td>17</td>
<td>37</td>
</tr>
<tr>
<td>ID24</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>ID25</td>
<td>3086</td>
<td>3770</td>
<td>2401</td>
<td>3462</td>
</tr>
<tr>
<td>ID26</td>
<td>5201</td>
<td>5892</td>
<td>4491</td>
<td>3414</td>
</tr>
<tr>
<td>ID27</td>
<td>2810</td>
<td>2802</td>
<td>2180</td>
<td>1881</td>
</tr>
<tr>
<td>ID28</td>
<td>193</td>
<td>227</td>
<td>333</td>
<td>221</td>
</tr>
<tr>
<td>ID29</td>
<td>341</td>
<td>301</td>
<td>178</td>
<td>219</td>
</tr>
<tr>
<td>ID30</td>
<td>218</td>
<td>205</td>
<td>187</td>
<td>160</td>
</tr>
<tr>
<td>ID31</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>ID32</td>
<td>310</td>
<td>635</td>
<td>295</td>
<td>262</td>
</tr>
<tr>
<td>ID33</td>
<td>12</td>
<td>10</td>
<td>17</td>
<td>25</td>
</tr>
<tr>
<td>ID34</td>
<td>736</td>
<td>318</td>
<td>226</td>
<td>559</td>
</tr>
<tr>
<td>ID35</td>
<td>3019</td>
<td>2678</td>
<td>2735</td>
<td>2283</td>
</tr>
<tr>
<td>ID36</td>
<td>2043</td>
<td>2112</td>
<td>2022</td>
<td>1560</td>
</tr>
<tr>
<td>ID37</td>
<td>3448</td>
<td>2435</td>
<td>1629</td>
<td>2206</td>
</tr>
<tr>
<td>ID38</td>
<td>3791</td>
<td>3213</td>
<td>3138</td>
<td>2729</td>
</tr>
<tr>
<td>ID39</td>
<td>5089</td>
<td>4662</td>
<td>4163</td>
<td>4248</td>
</tr>
<tr>
<td>ID40</td>
<td>145</td>
<td>63</td>
<td>46</td>
<td>83</td>
</tr>
<tr>
<td>ID41</td>
<td>644</td>
<td>590</td>
<td>1591</td>
<td>756</td>
</tr>
<tr>
<td>ID42</td>
<td>0</td>
<td>14</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

### Table B.1: Number of observations per news class in each yearly dataset for years 2012 - 2016.
## APPENDICES

<table>
<thead>
<tr>
<th>ID122</th>
<th>3</th>
<th>2</th>
<th>0</th>
<th>3</th>
<th>8</th>
<th>3</th>
<th>6</th>
<th>7</th>
<th>9</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID123</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>8</td>
<td>0</td>
<td>67</td>
<td>56</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>ID124</td>
<td>1176</td>
<td>175</td>
<td>35</td>
<td>98</td>
<td>5</td>
<td>12049</td>
<td>5851</td>
<td>4562</td>
<td>10126</td>
<td>14869</td>
</tr>
<tr>
<td>ID125</td>
<td>13</td>
<td>79</td>
<td>12</td>
<td>10</td>
<td>2930</td>
<td>225</td>
<td>854</td>
<td>314</td>
<td>346</td>
<td>2189</td>
</tr>
<tr>
<td>ID126</td>
<td>961</td>
<td>289</td>
<td>220</td>
<td>278</td>
<td>3875</td>
<td>16718</td>
<td>9876</td>
<td>8163</td>
<td>12655</td>
<td>23761</td>
</tr>
<tr>
<td>ID127</td>
<td>36</td>
<td>31</td>
<td>64</td>
<td>75</td>
<td>50</td>
<td>4610</td>
<td>4161</td>
<td>5475</td>
<td>6916</td>
<td>4764</td>
</tr>
<tr>
<td>ID128</td>
<td>14</td>
<td>30</td>
<td>26</td>
<td>33</td>
<td>8</td>
<td>3158</td>
<td>3222</td>
<td>3029</td>
<td>3433</td>
<td>3273</td>
</tr>
<tr>
<td>ID129</td>
<td>37</td>
<td>46</td>
<td>9</td>
<td>60</td>
<td>59</td>
<td>5137</td>
<td>6251</td>
<td>6242</td>
<td>6588</td>
<td>5706</td>
</tr>
<tr>
<td>ID130</td>
<td>1313</td>
<td>1273</td>
<td>1036</td>
<td>1048</td>
<td>563</td>
<td>9576</td>
<td>40072</td>
<td>48585</td>
<td>40092</td>
<td>45311</td>
</tr>
<tr>
<td>ID131</td>
<td>1996</td>
<td>2809</td>
<td>2494</td>
<td>2096</td>
<td>1428</td>
<td>36242</td>
<td>37439</td>
<td>37196</td>
<td>37227</td>
<td>30123</td>
</tr>
<tr>
<td>ID132</td>
<td>1105</td>
<td>12049</td>
<td>16335</td>
<td>15024</td>
<td>35584</td>
<td>355715</td>
<td>348340</td>
<td>318950</td>
<td>249815</td>
<td></td>
</tr>
<tr>
<td>ID133</td>
<td>43</td>
<td>81</td>
<td>14</td>
<td>26</td>
<td>31</td>
<td>3158</td>
<td>3222</td>
<td>3029</td>
<td>3433</td>
<td>3273</td>
</tr>
<tr>
<td>ID134</td>
<td>15</td>
<td>30</td>
<td>51</td>
<td>26</td>
<td>28</td>
<td>5476</td>
<td>5481</td>
<td>5331</td>
<td>6316</td>
<td>5769</td>
</tr>
<tr>
<td>ID135</td>
<td>36</td>
<td>31</td>
<td>54</td>
<td>75</td>
<td>50</td>
<td>4616</td>
<td>4161</td>
<td>5475</td>
<td>6916</td>
<td>4764</td>
</tr>
<tr>
<td>ID136</td>
<td>13</td>
<td>79</td>
<td>12</td>
<td>10</td>
<td>2930</td>
<td>225</td>
<td>854</td>
<td>314</td>
<td>346</td>
<td>2189</td>
</tr>
<tr>
<td>ID137</td>
<td>14</td>
<td>30</td>
<td>26</td>
<td>33</td>
<td>8</td>
<td>3158</td>
<td>3222</td>
<td>3029</td>
<td>3433</td>
<td>3273</td>
</tr>
<tr>
<td>ID138</td>
<td>15</td>
<td>30</td>
<td>51</td>
<td>26</td>
<td>28</td>
<td>5476</td>
<td>5481</td>
<td>5331</td>
<td>6316</td>
<td>5769</td>
</tr>
<tr>
<td>ID139</td>
<td>100</td>
<td>80</td>
<td>101</td>
<td>54</td>
<td>27</td>
<td>1562</td>
<td>1676</td>
<td>1653</td>
<td>1798</td>
<td>1399</td>
</tr>
<tr>
<td>ID140</td>
<td>43</td>
<td>81</td>
<td>14</td>
<td>26</td>
<td>31</td>
<td>3158</td>
<td>3222</td>
<td>3029</td>
<td>3433</td>
<td>3273</td>
</tr>
<tr>
<td>ID141</td>
<td>36</td>
<td>31</td>
<td>54</td>
<td>75</td>
<td>50</td>
<td>4616</td>
<td>4161</td>
<td>5475</td>
<td>6916</td>
<td>4764</td>
</tr>
<tr>
<td>ID142</td>
<td>2860</td>
<td>3041</td>
<td>2580</td>
<td>2827</td>
<td>3164</td>
<td>61599</td>
<td>61690</td>
<td>53175</td>
<td>61111</td>
<td>67311</td>
</tr>
<tr>
<td>ID143</td>
<td>3819</td>
<td>3656</td>
<td>3943</td>
<td>3671</td>
<td>3485</td>
<td>73910</td>
<td>65763</td>
<td>68093</td>
<td>74179</td>
<td>62365</td>
</tr>
<tr>
<td>ID144</td>
<td>3535</td>
<td>5335</td>
<td>4535</td>
<td>3735</td>
<td>3935</td>
<td>12955</td>
<td>120638</td>
<td>122299</td>
<td>121711</td>
<td>105996</td>
</tr>
<tr>
<td>ID145</td>
<td>223</td>
<td>352</td>
<td>240</td>
<td>464</td>
<td>507</td>
<td>15854</td>
<td>27779</td>
<td>20867</td>
<td>35468</td>
<td>46354</td>
</tr>
<tr>
<td>ID146</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>ID147</td>
<td>22</td>
<td>34</td>
<td>15</td>
<td>13</td>
<td>7</td>
<td>1018</td>
<td>1179</td>
<td>2110</td>
<td>4476</td>
<td>3049</td>
</tr>
<tr>
<td>ID148</td>
<td>30</td>
<td>2</td>
<td>46</td>
<td>4</td>
<td>5</td>
<td>2039</td>
<td>2114</td>
<td>2348</td>
<td>3370</td>
<td>5671</td>
</tr>
<tr>
<td>ID149</td>
<td>10253</td>
<td>11872</td>
<td>14766</td>
<td>14759</td>
<td>12664</td>
<td>99913</td>
<td>99380</td>
<td>116191</td>
<td>148306</td>
<td>153867</td>
</tr>
<tr>
<td>ID150</td>
<td>276</td>
<td>150</td>
<td>406</td>
<td>322</td>
<td>680</td>
<td>20316</td>
<td>20911</td>
<td>34580</td>
<td>44359</td>
<td>35897</td>
</tr>
<tr>
<td>ID151</td>
<td>3086</td>
<td>3775</td>
<td>2401</td>
<td>3462</td>
<td>1847</td>
<td>71557</td>
<td>73815</td>
<td>70336</td>
<td>72942</td>
<td>42131</td>
</tr>
<tr>
<td>ID152</td>
<td>5200</td>
<td>5891</td>
<td>4497</td>
<td>3422</td>
<td>3009</td>
<td>100241</td>
<td>112891</td>
<td>101112</td>
<td>96135</td>
<td>66804</td>
</tr>
<tr>
<td>ID153</td>
<td>2434</td>
<td>2444</td>
<td>2243</td>
<td>1869</td>
<td>1784</td>
<td>55120</td>
<td>50172</td>
<td>48309</td>
<td>47228</td>
<td>41546</td>
</tr>
<tr>
<td>ID154</td>
<td>3096</td>
<td>3784</td>
<td>2408</td>
<td>3463</td>
<td>1856</td>
<td>112585</td>
<td>108267</td>
<td>114397</td>
<td>132735</td>
<td>81660</td>
</tr>
<tr>
<td>ID155</td>
<td>5202</td>
<td>5892</td>
<td>4497</td>
<td>3422</td>
<td>3009</td>
<td>100313</td>
<td>112985</td>
<td>101112</td>
<td>96135</td>
<td>66804</td>
</tr>
<tr>
<td>ID156</td>
<td>2439</td>
<td>2463</td>
<td>2247</td>
<td>1872</td>
<td>1831</td>
<td>101500</td>
<td>102056</td>
<td>77190</td>
<td>76481</td>
<td>77776</td>
</tr>
<tr>
<td>ID157</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>290</td>
<td>4367</td>
<td>231</td>
<td>339</td>
<td>129</td>
</tr>
<tr>
<td>ID158</td>
<td>15</td>
<td>10</td>
<td>14</td>
<td>29</td>
<td>4</td>
<td>1139</td>
<td>1653</td>
<td>798</td>
<td>582</td>
<td>699</td>
</tr>
<tr>
<td>ID159</td>
<td>13</td>
<td>10</td>
<td>9</td>
<td>13</td>
<td>3</td>
<td>717</td>
<td>761</td>
<td>668</td>
<td>623</td>
<td>662</td>
</tr>
<tr>
<td>ID160</td>
<td>1405</td>
<td>5985</td>
<td>6644</td>
<td>6419</td>
<td>5310</td>
<td>175883</td>
<td>218379</td>
<td>235606</td>
<td>221178</td>
<td>161198</td>
</tr>
<tr>
<td>ID161</td>
<td>1505</td>
<td>1923</td>
<td>1566</td>
<td>1292</td>
<td>1835</td>
<td>87623</td>
<td>109410</td>
<td>85960</td>
<td>80885</td>
<td>79693</td>
</tr>
<tr>
<td>ID162</td>
<td>497</td>
<td>6359</td>
<td>6850</td>
<td>6666</td>
<td>5212</td>
<td>218180</td>
<td>250524</td>
<td>270849</td>
<td>260064</td>
<td>196341</td>
</tr>
<tr>
<td>ID163</td>
<td>100</td>
<td>34</td>
<td>145</td>
<td>111</td>
<td>58</td>
<td>2621</td>
<td>3056</td>
<td>5587</td>
<td>7977</td>
<td>6257</td>
</tr>
<tr>
<td>ID164</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>24</td>
<td>56</td>
<td>23</td>
</tr>
<tr>
<td>ID165</td>
<td>125</td>
<td>286</td>
<td>719</td>
<td>441</td>
<td>448</td>
<td>83071</td>
<td>89338</td>
<td>126057</td>
<td>609890</td>
<td>158820</td>
</tr>
<tr>
<td>ID166</td>
<td>30</td>
<td>10</td>
<td>23</td>
<td>19</td>
<td>147</td>
<td>4990</td>
<td>6996</td>
<td>11664</td>
<td>11364</td>
<td>15860</td>
</tr>
</tbody>
</table>

81
## Appendix C

### Observations per News Class for Monthly Datasets

Table C.1: Number of observations per news class in each monthly dataset for months Jan - May 2015.

<table>
<thead>
<tr>
<th>ID</th>
<th>Canada Data</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>World Data</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jan</td>
<td>Feb</td>
<td>Mar</td>
<td>Apr</td>
<td>May</td>
<td>Jan</td>
<td>Feb</td>
<td>Mar</td>
<td>Apr</td>
<td>May</td>
</tr>
<tr>
<td>ID1</td>
<td>55</td>
<td>16</td>
<td>130</td>
<td>9</td>
<td>3</td>
<td>404</td>
<td>458</td>
<td>547</td>
<td>204</td>
<td>360</td>
</tr>
<tr>
<td>ID2</td>
<td>72</td>
<td>159</td>
<td>131</td>
<td>111</td>
<td>128</td>
<td>550</td>
<td>256</td>
<td>313</td>
<td>249</td>
<td>313</td>
</tr>
<tr>
<td>ID3</td>
<td>1848</td>
<td>1623</td>
<td>1679</td>
<td>1483</td>
<td>1525</td>
<td>16919</td>
<td>1947</td>
<td>2167</td>
<td>2197</td>
<td>21853</td>
</tr>
<tr>
<td>ID4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>5</td>
<td>12</td>
<td>2</td>
<td>70</td>
</tr>
<tr>
<td>ID5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ID6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>96</td>
<td>178</td>
<td>144</td>
<td>493</td>
<td>156</td>
</tr>
<tr>
<td>ID7</td>
<td>1924</td>
<td>1585</td>
<td>227</td>
<td>149</td>
<td>1496</td>
<td>4890</td>
<td>4447</td>
<td>3665</td>
<td>4207</td>
<td>3304</td>
</tr>
<tr>
<td>ID8</td>
<td>226</td>
<td>83</td>
<td>151</td>
<td>99</td>
<td>121</td>
<td>2505</td>
<td>2661</td>
<td>3813</td>
<td>3465</td>
<td>3572</td>
</tr>
<tr>
<td>ID9</td>
<td>205</td>
<td>167</td>
<td>196</td>
<td>239</td>
<td>210</td>
<td>3775</td>
<td>3254</td>
<td>4168</td>
<td>4582</td>
<td>4796</td>
</tr>
<tr>
<td>ID10</td>
<td>215</td>
<td>100</td>
<td>118</td>
<td>169</td>
<td>92</td>
<td>2264</td>
<td>2684</td>
<td>1976</td>
<td>1919</td>
<td>1643</td>
</tr>
<tr>
<td>ID11</td>
<td>111</td>
<td>132</td>
<td>137</td>
<td>130</td>
<td>91</td>
<td>1211</td>
<td>829</td>
<td>1080</td>
<td>548</td>
<td>948</td>
</tr>
<tr>
<td>ID12</td>
<td>115</td>
<td>194</td>
<td>236</td>
<td>178</td>
<td>75</td>
<td>3944</td>
<td>4106</td>
<td>5642</td>
<td>3974</td>
<td>4169</td>
</tr>
<tr>
<td>ID13</td>
<td>25</td>
<td>33</td>
<td>52</td>
<td>21</td>
<td>58</td>
<td>828</td>
<td>1026</td>
<td>792</td>
<td>1164</td>
<td>1064</td>
</tr>
<tr>
<td>ID14</td>
<td>12</td>
<td>21</td>
<td>50</td>
<td>13</td>
<td>13</td>
<td>750</td>
<td>500</td>
<td>816</td>
<td>555</td>
<td>441</td>
</tr>
<tr>
<td>ID15</td>
<td>11</td>
<td>9</td>
<td>10</td>
<td>57</td>
<td>32</td>
<td>1289</td>
<td>1294</td>
<td>1267</td>
<td>1366</td>
<td>1033</td>
</tr>
<tr>
<td>ID16</td>
<td>5</td>
<td>33</td>
<td>12</td>
<td>30</td>
<td>9</td>
<td>689</td>
<td>571</td>
<td>276</td>
<td>129</td>
<td>224</td>
</tr>
<tr>
<td>ID17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ID18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>11</td>
<td>2</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td>ID19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>11</td>
<td>3</td>
<td>14</td>
<td>61</td>
</tr>
<tr>
<td>ID20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>ID21</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>59</td>
<td>82</td>
<td>73</td>
<td>37</td>
<td>50</td>
</tr>
<tr>
<td>ID22</td>
<td>162</td>
<td>24</td>
<td>56</td>
<td>41</td>
<td>21</td>
<td>3940</td>
<td>2289</td>
<td>2728</td>
<td>3068</td>
<td>3089</td>
</tr>
<tr>
<td>ID23</td>
<td>0</td>
<td>22</td>
<td>9</td>
<td>4</td>
<td>2</td>
<td>441</td>
<td>389</td>
<td>371</td>
<td>527</td>
<td>392</td>
</tr>
<tr>
<td>ID24</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>ID25</td>
<td>27</td>
<td>178</td>
<td>351</td>
<td>146</td>
<td>241</td>
<td>15592</td>
<td>9792</td>
<td>12914</td>
<td>9390</td>
<td>9313</td>
</tr>
<tr>
<td>ID26</td>
<td>300</td>
<td>284</td>
<td>285</td>
<td>329</td>
<td>283</td>
<td>8138</td>
<td>7456</td>
<td>8368</td>
<td>9209</td>
<td>8874</td>
</tr>
<tr>
<td>ID27</td>
<td>218</td>
<td>163</td>
<td>130</td>
<td>154</td>
<td>192</td>
<td>9871</td>
<td>7334</td>
<td>6882</td>
<td>7621</td>
<td>6129</td>
</tr>
<tr>
<td>ID28</td>
<td>0</td>
<td>23</td>
<td>45</td>
<td>45</td>
<td>10</td>
<td>1187</td>
<td>870</td>
<td>1292</td>
<td>939</td>
<td>922</td>
</tr>
<tr>
<td>ID29</td>
<td>352</td>
<td>58</td>
<td>33</td>
<td>24</td>
<td>44</td>
<td>1689</td>
<td>1029</td>
<td>1472</td>
<td>1169</td>
<td>1425</td>
</tr>
<tr>
<td>ID30</td>
<td>42</td>
<td>9</td>
<td>13</td>
<td>29</td>
<td>20</td>
<td>1241</td>
<td>962</td>
<td>319</td>
<td>675</td>
<td>675</td>
</tr>
<tr>
<td>ID31</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ID32</td>
<td>11</td>
<td>4</td>
<td>13</td>
<td>56</td>
<td>12</td>
<td>99</td>
<td>102</td>
<td>182</td>
<td>291</td>
<td>360</td>
</tr>
<tr>
<td>ID33</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>13</td>
<td>11</td>
<td>48</td>
<td>2</td>
</tr>
<tr>
<td>ID34</td>
<td>82</td>
<td>59</td>
<td>52</td>
<td>66</td>
<td>25</td>
<td>1274</td>
<td>1550</td>
<td>1013</td>
<td>1426</td>
<td>1135</td>
</tr>
<tr>
<td>ID35</td>
<td>364</td>
<td>158</td>
<td>317</td>
<td>219</td>
<td>171</td>
<td>3740</td>
<td>6002</td>
<td>5921</td>
<td>4908</td>
<td>5460</td>
</tr>
<tr>
<td>ID36</td>
<td>180</td>
<td>184</td>
<td>210</td>
<td>105</td>
<td>126</td>
<td>2828</td>
<td>1921</td>
<td>3961</td>
<td>2123</td>
<td>2138</td>
</tr>
<tr>
<td>ID37</td>
<td>302</td>
<td>407</td>
<td>158</td>
<td>237</td>
<td>158</td>
<td>5029</td>
<td>4707</td>
<td>3830</td>
<td>4375</td>
<td>3754</td>
</tr>
<tr>
<td>ID38</td>
<td>402</td>
<td>198</td>
<td>351</td>
<td>269</td>
<td>223</td>
<td>4284</td>
<td>5460</td>
<td>6571</td>
<td>5999</td>
<td>5993</td>
</tr>
<tr>
<td>ID39</td>
<td>324</td>
<td>281</td>
<td>391</td>
<td>193</td>
<td>322</td>
<td>5073</td>
<td>4521</td>
<td>8148</td>
<td>5927</td>
<td>6036</td>
</tr>
<tr>
<td>ID40</td>
<td>14</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>1810</td>
<td>1822</td>
<td>770</td>
<td>946</td>
<td>1156</td>
</tr>
<tr>
<td>ID41</td>
<td>161</td>
<td>107</td>
<td>50</td>
<td>22</td>
<td>181</td>
<td>5864</td>
<td>4810</td>
<td>4981</td>
<td>1248</td>
<td>4703</td>
</tr>
<tr>
<td>ID42</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>386</td>
<td>117</td>
<td>144</td>
<td>300</td>
<td>118</td>
</tr>
<tr>
<td>ID</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>ID123</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>ID124</td>
<td>13</td>
<td>10</td>
<td>15</td>
<td>14</td>
<td>16</td>
<td>15</td>
<td>15</td>
<td>14</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>ID125</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>33</td>
<td>35</td>
<td>37</td>
<td>55</td>
<td>29</td>
</tr>
<tr>
<td>ID126</td>
<td>18</td>
<td>41</td>
<td>25</td>
<td>25</td>
<td>34</td>
<td>1148</td>
<td>1188</td>
<td>1440</td>
<td>1435</td>
<td>1124</td>
</tr>
<tr>
<td>ID127</td>
<td>10</td>
<td>16</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>411</td>
<td>417</td>
<td>574</td>
<td>547</td>
<td>595</td>
</tr>
<tr>
<td>ID128</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>269</td>
<td>269</td>
<td>230</td>
<td>320</td>
<td>296</td>
<td></td>
</tr>
<tr>
<td>ID129</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>760</td>
<td>701</td>
<td>775</td>
<td>499</td>
<td>513</td>
</tr>
<tr>
<td>ID130</td>
<td>251</td>
<td>170</td>
<td>56</td>
<td>85</td>
<td>55</td>
<td>3925</td>
<td>2674</td>
<td>3411</td>
<td>3831</td>
<td>2912</td>
</tr>
<tr>
<td>ID131</td>
<td>303</td>
<td>128</td>
<td>119</td>
<td>231</td>
<td>284</td>
<td>8021</td>
<td>2545</td>
<td>4768</td>
<td>3829</td>
<td>3006</td>
</tr>
<tr>
<td>ID132</td>
<td>1319</td>
<td>1512</td>
<td>1911</td>
<td>1854</td>
<td>1342</td>
<td>26227</td>
<td>24000</td>
<td>30309</td>
<td>27013</td>
<td>21813</td>
</tr>
<tr>
<td>ID133</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>197</td>
<td>174</td>
<td>138</td>
<td>176</td>
<td>82</td>
</tr>
<tr>
<td>ID134</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>16</td>
<td>2</td>
<td>357</td>
<td>817</td>
<td>136</td>
<td>660</td>
<td>591</td>
</tr>
<tr>
<td>ID135</td>
<td>7</td>
<td>14</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>155</td>
<td>141</td>
<td>129</td>
<td>196</td>
<td>90</td>
</tr>
<tr>
<td>ID136</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>177</td>
<td>173</td>
<td>138</td>
<td>176</td>
<td>82</td>
</tr>
<tr>
<td>ID137</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>16</td>
<td>2</td>
<td>357</td>
<td>817</td>
<td>136</td>
<td>660</td>
<td>591</td>
</tr>
<tr>
<td>ID138</td>
<td>7</td>
<td>14</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>155</td>
<td>141</td>
<td>129</td>
<td>196</td>
<td>90</td>
</tr>
<tr>
<td>ID139</td>
<td>3</td>
<td>9</td>
<td>24</td>
<td>19</td>
<td>1</td>
<td>412</td>
<td>889</td>
<td>616</td>
<td>768</td>
<td>730</td>
</tr>
<tr>
<td>ID140</td>
<td>2</td>
<td>17</td>
<td>18</td>
<td>27</td>
<td>8</td>
<td>236</td>
<td>587</td>
<td>416</td>
<td>626</td>
<td>490</td>
</tr>
<tr>
<td>ID141</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>361</td>
<td>11</td>
<td>52</td>
<td>46</td>
<td>16</td>
</tr>
<tr>
<td>ID142</td>
<td>10</td>
<td>302</td>
<td>410</td>
<td>164</td>
<td>964</td>
<td>3806</td>
<td>8658</td>
<td>3686</td>
<td>6886</td>
<td>8026</td>
</tr>
<tr>
<td>ID143</td>
<td>292</td>
<td>352</td>
<td>382</td>
<td>299</td>
<td>204</td>
<td>6717</td>
<td>8382</td>
<td>4501</td>
<td>7091</td>
<td>7159</td>
</tr>
<tr>
<td>ID144</td>
<td>355</td>
<td>596</td>
<td>393</td>
<td>612</td>
<td>10564</td>
<td>16066</td>
<td>8548</td>
<td>12176</td>
<td>14552</td>
<td></td>
</tr>
<tr>
<td>ID145</td>
<td>3</td>
<td>7</td>
<td>17</td>
<td>35</td>
<td>4</td>
<td>2828</td>
<td>2516</td>
<td>2140</td>
<td>2117</td>
<td>1449</td>
</tr>
<tr>
<td>ID146</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ID147</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>307</td>
<td>81</td>
<td>224</td>
<td>399</td>
<td>194</td>
</tr>
<tr>
<td>ID148</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>100</td>
<td>379</td>
<td>638</td>
<td>797</td>
<td>511</td>
</tr>
<tr>
<td>ID149</td>
<td>907</td>
<td>1346</td>
<td>958</td>
<td>1723</td>
<td>1778</td>
<td>14235</td>
<td>11539</td>
<td>230</td>
<td>20104</td>
<td>9947</td>
</tr>
<tr>
<td>ID150</td>
<td>2</td>
<td>19</td>
<td>12</td>
<td>150</td>
<td>55</td>
<td>3518</td>
<td>3282</td>
<td>3844</td>
<td>3881</td>
<td>4039</td>
</tr>
<tr>
<td>ID151</td>
<td>272</td>
<td>178</td>
<td>347</td>
<td>145</td>
<td>245</td>
<td>6128</td>
<td>5128</td>
<td>8062</td>
<td>6188</td>
<td>6722</td>
</tr>
<tr>
<td>ID152</td>
<td>750</td>
<td>284</td>
<td>283</td>
<td>329</td>
<td>283</td>
<td>7583</td>
<td>7107</td>
<td>7973</td>
<td>8944</td>
<td>8479</td>
</tr>
<tr>
<td>ID153</td>
<td>226</td>
<td>163</td>
<td>135</td>
<td>159</td>
<td>179</td>
<td>4174</td>
<td>4391</td>
<td>4310</td>
<td>4731</td>
<td>4110</td>
</tr>
<tr>
<td>ID154</td>
<td>273</td>
<td>178</td>
<td>351</td>
<td>146</td>
<td>245</td>
<td>15511</td>
<td>9806</td>
<td>12909</td>
<td>9314</td>
<td>9326</td>
</tr>
<tr>
<td>ID155</td>
<td>800</td>
<td>284</td>
<td>283</td>
<td>329</td>
<td>283</td>
<td>8147</td>
<td>7462</td>
<td>8174</td>
<td>9216</td>
<td>8882</td>
</tr>
<tr>
<td>ID156</td>
<td>225</td>
<td>164</td>
<td>135</td>
<td>159</td>
<td>200</td>
<td>6550</td>
<td>7482</td>
<td>7141</td>
<td>7990</td>
<td>6918</td>
</tr>
<tr>
<td>ID157</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td>36</td>
<td>44</td>
<td>46</td>
<td>18</td>
</tr>
<tr>
<td>ID158</td>
<td>1</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>58</td>
<td>58</td>
<td>63</td>
<td>39</td>
<td>49</td>
</tr>
<tr>
<td>ID159</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>111</td>
<td>81</td>
<td>44</td>
<td>73</td>
<td>150</td>
</tr>
<tr>
<td>ID160</td>
<td>568</td>
<td>514</td>
<td>604</td>
<td>517</td>
<td>557</td>
<td>18712</td>
<td>18583</td>
<td>20668</td>
<td>20038</td>
<td>10371</td>
</tr>
<tr>
<td>ID161</td>
<td>84</td>
<td>151</td>
<td>159</td>
<td>163</td>
<td>177</td>
<td>6438</td>
<td>7205</td>
<td>10818</td>
<td>9584</td>
<td>8422</td>
</tr>
<tr>
<td>ID162</td>
<td>564</td>
<td>624</td>
<td>575</td>
<td>632</td>
<td>548</td>
<td>21990</td>
<td>22726</td>
<td>23637</td>
<td>23081</td>
<td>21588</td>
</tr>
<tr>
<td>ID163</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>6</td>
<td>11</td>
<td>568</td>
<td>331</td>
<td>560</td>
<td>615</td>
<td>346</td>
</tr>
<tr>
<td>ID164</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ID165</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ID166</td>
<td>74</td>
<td>13</td>
<td>75</td>
<td>95</td>
<td>3</td>
<td>13769</td>
<td>12066</td>
<td>13114</td>
<td>12096</td>
<td>12726</td>
</tr>
<tr>
<td>ID167</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>9</td>
<td>127</td>
<td>105</td>
<td>174</td>
<td>246</td>
<td>291</td>
</tr>
<tr>
<td>ID168</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>778</td>
<td>1494</td>
<td>611</td>
<td>2114</td>
<td>1159</td>
</tr>
</tbody>
</table>
Appendix D

Observations per News Class for Daily Datasets

*Table D.1: Number of observations per news class in each daily dataset for dates 1st - 7th of March, 2015.*

<table>
<thead>
<tr>
<th>ID</th>
<th>Canada Data</th>
<th></th>
<th>World Data</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID1</td>
<td>0 0 0 10 0 0 0</td>
<td>0 0 15 11 0 7 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID2</td>
<td>0 11 3 1 4 2 0</td>
<td>5 59 102 101 68 127 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID3</td>
<td>2 23 46 31 27 59 2</td>
<td>71 273 1096 573 1119 363 38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID4</td>
<td>0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID5</td>
<td>0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID6</td>
<td>0 0 0 0 0 0 0</td>
<td>4 0 0 2 2 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID7</td>
<td>0 23 5 7 12 15 0</td>
<td>6 217 264 107 174 155 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID8</td>
<td>0 6 3 0 1 4 0</td>
<td>0 120 249 214 199 169 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID9</td>
<td>0 17 2 14 11 13 0</td>
<td>5 170 152 169 182 259 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID10</td>
<td>0 12 3 0 0 0 0</td>
<td>11 158 22 49 72 28 12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID11</td>
<td>0 0 3 11 0 0 0</td>
<td>3 156 1 16 2 7 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID12</td>
<td>0 1 0 0 2 4 0</td>
<td>15 162 280 276 229 306 20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID13</td>
<td>0 15 0 0 3 2 1</td>
<td>35 59 24 19 24 65 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID14</td>
<td>0 0 3 1 0 4 0</td>
<td>3 20 50 6 13 99 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID15</td>
<td>0 2 0 0 0 3 0</td>
<td>23 49 29 30 23 91 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID16</td>
<td>0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID17</td>
<td>0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID18</td>
<td>0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID19</td>
<td>0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID20</td>
<td>0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID21</td>
<td>0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID22</td>
<td>1 1 3 2 0 1 0</td>
<td>209 139 69 26 27 45 49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID23</td>
<td>0 0 0 0 0 0 0</td>
<td>4 11 11 4 8 24 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID24</td>
<td>0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID25</td>
<td>1 11 14 13 10 23 0</td>
<td>17 514 542 564 400 557 39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID26</td>
<td>0 10 19 6 14 17 0</td>
<td>4 441 379 290 442 274 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID27</td>
<td>0 2 6 9 5 9 4 0</td>
<td>2 162 352 224 269 218 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID28</td>
<td>0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID29</td>
<td>0 0 0 0 0 0 0</td>
<td>1 88 87 32 16 37 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID30</td>
<td>0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID31</td>
<td>0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID32</td>
<td>0 0 0 13 0 0 0</td>
<td>0 8 0 14 5 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID33</td>
<td>0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID34</td>
<td>19 0 0 1 2 0 0</td>
<td>2 52 73 34 70 56 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID35</td>
<td>0 4 8 0 1 4 0</td>
<td>2 234 349 392 326 229 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID36</td>
<td>1 14 3 0 0 0 0</td>
<td>5 101 104 30 324 173 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID37</td>
<td>0 23 5 7 12 15 0</td>
<td>6 233 222 173 180 174 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID38</td>
<td>0 6 9 0 1 4 0</td>
<td>2 253 187 418 345 252 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID39</td>
<td>0 0 0 1 13 13 0</td>
<td>2 432 231 216 355 108 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID40</td>
<td>1 0 0 0 0 0 0</td>
<td>14 23 12 17 43 8 55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID41</td>
<td>0 0 1 0 1 1 9</td>
<td>84 146 106 155 60 165 310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID42</td>
<td>0 0 0 0 0 0 0</td>
<td>5 12 2 0 16 13 25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## APPENDICES

<table>
<thead>
<tr>
<th>ID100</th>
<th>0</th>
<th>1</th>
<th>3</th>
<th>0</th>
<th>12</th>
<th>6</th>
<th>0</th>
<th>6</th>
<th>11</th>
<th>44</th>
<th>71</th>
<th>60</th>
<th>64</th>
<th>64</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID101</td>
<td>0</td>
<td>73</td>
<td>45</td>
<td>30</td>
<td>85</td>
<td>65</td>
<td>0</td>
<td>4</td>
<td>364</td>
<td>389</td>
<td>359</td>
<td>466</td>
<td>341</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>ID102</td>
<td>0</td>
<td>5</td>
<td>19</td>
<td>12</td>
<td>29</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>80</td>
<td>179</td>
<td>202</td>
<td>156</td>
<td>115</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>ID103</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>46</td>
<td>16</td>
<td>30</td>
<td>146</td>
<td>8</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ID104</td>
<td>0</td>
<td>1</td>
<td>21</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>8</td>
<td>38</td>
<td>52</td>
<td>100</td>
<td>116</td>
<td>41</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>ID105</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>29</td>
<td>110</td>
<td>36</td>
<td>143</td>
<td>31</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>ID106</td>
<td>0</td>
<td>33</td>
<td>121</td>
<td>312</td>
<td>373</td>
<td>168</td>
<td>6</td>
<td>98</td>
<td>1509</td>
<td>1942</td>
<td>3011</td>
<td>2646</td>
<td>1702</td>
<td>493</td>
<td></td>
</tr>
<tr>
<td>ID107</td>
<td>0</td>
<td>168</td>
<td>185</td>
<td>248</td>
<td>360</td>
<td>146</td>
<td>0</td>
<td>10</td>
<td>1690</td>
<td>1447</td>
<td>1946</td>
<td>1965</td>
<td>892</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>ID108</td>
<td>0</td>
<td>61</td>
<td>130</td>
<td>320</td>
<td>390</td>
<td>148</td>
<td>2</td>
<td>70</td>
<td>1312</td>
<td>2763</td>
<td>4103</td>
<td>3006</td>
<td>2184</td>
<td>357</td>
<td></td>
</tr>
<tr>
<td>ID109</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ID110</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ID111</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ID112</td>
<td>0</td>
<td>11</td>
<td>25</td>
<td>4</td>
<td>32</td>
<td>25</td>
<td>0</td>
<td>15</td>
<td>67</td>
<td>11</td>
<td>49</td>
<td>36</td>
<td>28</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ID113</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ID114</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ID115</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ID116</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ID117</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ID118</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ID119</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

86