Three Essays on Business Analytics:

time-series causality, panel data analysis, and design of experiments

by

Yaser Abolghsemi Dehaghani

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Management Sciences

Waterloo, Ontario, Canada, 2020

© Yaser Abolghasemi Dehaghani 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Chunming (Victor) Shi Associate Professor, Lazaridis School of Business and Economics, Wilfrid Laurier University
Supervisor:	Stanko Dimitrov Associate Professor, Dept. of Management Sciences, University of Waterloo
Internal Members:	Qi-Ming He Professor, Dept. of Management Sciences, University of Waterloo
	Brian Paul Cozzarin Associate Professor, Dept. of Management Sciences, University of Waterloo
Internal-External Member:	Greg Rice Assistant Professor, Dept. of Statistics and Actuarial Science, University of Waterloo

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This dissertation includes three empirical studies focusing on the applications of Business Analytics in the context of financial markets and the online advertisement industry. The first essay examines the existence of a causal relationship between various prediction markets and global financial markets time series. This essay uses over 27 different countries and regions' financial market data (Dow Jones Global Indexes) and uses the Toda-Yamamoto causality test. Preliminary results indicate that prediction markets may be used to predict some global financial markets. From a managerial perspective, our result quantifies the connection between some countries' economy, as measured by a financial index, and the political events captured by the prediction markets we consider.

The next two essays focus on the online advertising industry's business policies. The second essay uses a panel data analysis to compare the effect of two different IP protection policies, Monetize and Track, on YouTube music channels' viewership. This research provides insights for content owners on how IP protection policies on user-generated contents (UGCs) affect their YouTube channel viewership, and on how UGCs impact their ability to maximize profit. The third and final essay proposes a new data-driven statistical framework (DDSF) to determine what ad formats maximize a company's revenue generated from online advertising. The developed DDSF is applied in a real-world experiment. The experiment results help our YouTube industry partner determine what ad formats to run on their videos in order to trade off two key performance indicators of interest.

Acknowledgements

In the Name of Allah, the Most Beneficent, the Most Merciful

All the praises and thanks be to Allah, the Lord of the Universe

First and foremost, I would like to thank Professor Stanko Dimitrov for his fantastic supervision during my Ph.D. program. This thesis is the outcome of hours of fruitful discussion with him. He taught me how to deal with new and complex problems rigorously and formally. Prof. Dimotrov's immense knowledge and integrity were always inspiring to me.

Second, I am grateful to Professor Chunming (Victor) Shi from the Wilfrid Laurier University for insightful evaluation my dissertation as an external examiner. I would like to express my gratitude to my advisory committee members Professors Qi-Ming He, Brian Paul Cozzarin, and Greg Rice for their comments and suggestions.

I would also like to thank our industry partners, Cody Allard and Ari Paunonen, for giving us the chance to get access to their business datasets and helping us to solve a real-world problem. Ari and Cody's business insights and technical comments were always fantastic. I've been blessed to work with wonderful partners like them.

Finally, the most sincere appreciation goes to my parents, Mohammadjavad and Zahra, for their unconditional love and support in my entire life. Last but not least, I would like to thank my dearest wife, Hamideh, for her selfless support and excellent companionship.

Dedication

To my wonderful wife, ***Hamideh***,

and our lovely sons, ***Ali*** & ***Mahdi***.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Business Analytics (BA) integrates data and analytic systems with business requirements to determine and leverage business opportunities. BA uses data, computer machines, and statistics to understand, manage, and solve business problems. In other words, BA is the science of making better business decisions. The focus of this dissertation is not on BA theories. Instead, this dissertation focuses on the applications of BA and uses various statistical methods in its three different essays.

In this dissertation, each essay discusses different topics in BA in the area of financial markets and online advertising. The first essay, which is independent of the other two essays, investigates the connection of prediction markets and various global financial markets; the other two essays address an industry-partner's business problems in the online advertising industry. Now we present an overview for each essay:

- **Essay One:** *Determining the causality between U.S. presidential prediction markets and global financial markets.*

  Prediction markets trade securities with final prices contingent on the outcome of future events, for example who will win the next political election. We show how the outcome of a United States presidential election, information captured by prediction markets, impacts global financial markets. We investigate the existence of a causal relationship between various prediction markets and global financial markets time series for over 27 different countries and regions using Dow Jones Global Indexes. We construct vector auto-regressive models and use the Toad-Yamamoto causality test to deal with non-stationary time series. Preliminary results indicate that prediction markets may be used to predict some global financial markets.

- **Essay Two:** *Determining the impact of different IP protection policies of YouTube videos on YouTube channels.*

  The YouTube IP protection system (i.e, Content ID system) contains more than 50 million active reference files in its database. It helps YouTube partners to make claims for the use of hundreds of millions of infringing videos on YouTube by blocking, tracking, or monetizing these user-generated contents (UGCs). Monetizing UGCs generated more than 3 billion dollars in ad revenue for content creators up to the end of 2017.

  In this research, we use panel data analysis to compare the effect of two different IP protection policies, Monetize and Track, on YouTube channel viewership. This research provides insights for content owners on how IP protection policies on UGCs affect their YouTube channel viewership, and on how UGCs impact their ability to maximize profit. Additionally, the results of this study provide insights on the impact of a copyright holder making his/her intellectual property openly available and how this decision influences the copyright holders' business.

- **Essay Three:** *A data-driven statistical framework for maximizing business revenue generating processes: An application to advertising policies on YouTube.*

  Internet advertising revenues in the United States was more than 107 billion for the year 2018 that was increased by 21.8% over 2017. A portion of this advertising revenue is earned by companies that monetize their content (i.e, display various ad formats) on different online advertising platforms. In the first half of this research a new data-driven statistical framework (DDSF) is proposed to determine what ad formats maximize a company's revenue. The DDSF is based on a quasi-experimental design that uses a Genetic Algorithm for treatment assignment. The proposed DDSF has four stages 1) Design of Experiment, 2) Data Collection, 3) Running Experiment, and 4) Making Decision. The statistical framework performs non-parametric analysis, the Kruskal-Wallis and Wilcoxon post-hot test. The DDSF also presents the Pareto efficiency charts to help decision makers determine the best treatment groups.

In the second half of this research the developed DDSF is applied in a real-world experiment. The experiment results help a YouTube industry partner determine what ad formats to run on their videos in order to maximize two key performance indicators of interest. The key performance indicators of interest to the industry partner are channel's average viewership and gross revenue.

The remainder of this dissertation is divided into three chapters. Essay 1 is found in Chapter 2. Essays 2 and 3 are found in Chapters 3 and 4, respectively. We conclude the dissertation after the conclusion of Chapter 4.

Chapter 2

Determining the causality between U.S. presidential prediction markets and global financial markets

2.1 Introduction

It is always a debate over what forces are shaping the global economy and finance. Recently, World Economic Forum researchers highlighted the following six factors as influences on the global economy in 2016: political populism, refugee crisis, Brexit, Russia's role in the world, weak growth and China's reforms (Borge, 2015). However, it is obvious that most of these factors have a common root which is politics that influence the global economy.

Globalization, as a process of international integration, is widely supported by countries to facilitate global transactions. Despite the growing integration of financial markets, which is a sign of economic globalization, a new government's policies might contradict with globalization fundamentals. For instance, the ideology of Nationalism, found in various western democracies, affects global financial markets, and may be one cause of the U.S.-China trade war costing global GDP as much as 600 billion dollars (Holland and Sam, 2019).

The United States (U.S.) has the world's largest economy, forming 24% of global gross domestic product (GDP), 20% of global foreign direct investment (FDI), and more 44% of global stock market capitalization (IMF, 2018). The U.S. stock market value in 2018 was around 30 trillion dollars, which is 5 times bigger than its closest competitor, which is China with 6.3 trillion dollars market capitalization (Hotten, 2019).

Considering the global influence of the U.S. economy, a growth surge in the U.S. economy could lead to a global economic boost. In contrast, uncertainty about the direction of U.S. policies could have the opposite effect. The U.S. presidential election has a significant effect on U.S. financial market performance. In every U.S. presidential election, since 1833, the Dow Jones Industrial Average (DJIA) increased 10.4% one year before the U.S. presidential elections and 6% throughout the election year; and the only exception to the found trend is the 2008 U.S. presidential election during which time the DJIA decreased 34% (Smith, 2016). Nickles (2004) explains that government administrators wanted to engage in practices, such as creating changes in taxation, that would influence fiscal policies just before a presidential election, which would in turn have a positive effect on stock prices, making investors assume that favourable conditions were ahead for businesses.

The U.S. financial markets' fluctuations caused by presidential elections will affect global financial markets indices. For example, Tokic (2003) investigate the interaction between the U.S. stock market and global financial market, and show a long-run relationship between the U.S. market and Australia, Japan, Hong Kong, New Zealand and Singapore financial markets. In this research we investigate the relationship between the U.S. presidential elections and global financial market indices.

Based on The Gordon Growth Model (GGM), a variant of a dividend discount model (DDM), the intrinsic value of a stock is based on a future series of dividends that grow at a constant rate (Gordon, 1962). Thus,stock prices correspond to the discounted sum of expected dividends, which is a measure for future economic performance, and any political event (policy or tax) that will benefit the market (its traded firms, stock owners, or market itself) is expected to increase stock prices. Similarly, stock prices are expected to decrease with political events that will harm the market. So, stock price movements are informative as they reflect changes in expectations of future economic performance. By definition, a financial market index change captures changes in the overall market. For a country financial market, the country financial market index is a leading indicator that helps to measure country's national economic health (Aylward and Glen, 2000; Harvey, 1989; Comincioli, 1996).

Indeed, U.S. presidential elections impact the U.S. economy, which is an "engine" of the world economy Arora and Vamvakidis (2004). In addition to impacting domestic policies, the party in the oval office determines U.S. policies towards other countries. Considering the fact that a country's well being is measured by its stock market financial index, we determine what is the impact of a U.S. presidential election on other countries' economy, as measured by financial market indices.

Considering the U.S. economic significance, which is the world's largest economy, the U.S. presidential election is always a key recurring political event that has profound consequences for the global economy. From this perspective, studies that help global economies and financial markets learn existence and direction of relationships between U.S. presidential elections and financial markets can be help to understand future global economic challenges.

We investigate the existence and direction of a causal relationship between the U.S. presidential election outcome and various global financial markets, using global financial indices. Financial indices are among the most important indicators that directly reflect countries or regions' financial health and prospects, and the U.S. presidential election prediction market in our analysis.

Prediction markets are futures markets in which securities are traded with prices contingent on the outcome of future events. The price of prediction market securities at any point in time yields a probabilistic estimate on the likelihood of the contingent event occurring at that point in time (Berg et al., 2003). One of the first electronic prediction markets, the Iowa Electronic Markets, started in 1988 to predict election outcomes in the United States of America (U.S.). Since their inception, prediction markets have grown in popularity due to their accuracy; prediction markets are shown to be more accurate than survey-based and forecasting approaches and can also address compensation problems in traditional forecasting tools by allowing participants to trade securities (Ho and Chen, 2007). Not only are prediction markets more accurate, they also aggregate information quicker than classical methods (Berg et al., 2008). Therefore, prediction markets are used for constructing robust decision support systems and predictive models (Berg and Rietz, 2003).

This research is motivated by the general quest to determine what and by how much U.S. presidential elections impact world financial and economic systems. Since, prediction market prices provide a point estimate on the probability of an event, U.S. presidential election in our case, occurring at any one point in time, putting the points together generates a time series of probability estimates. Similarly, world economic and financial data are also measured on a second-by-second or minute-by-minute time increment, for example the DOW Jones market indexes.

In particular, we determine the impact US presidential elections have on the global economy as measured by their impact on specific countries and regions financial markets. We use the Toda-Yamamoto (TY) procedure to determine the statistical causal relationship between prediction markets and global financial markets series.

Using the Toda-Yamamoto (TY) procedure, we determine the existence of causal relationships between the prediction-market and global-financial-markets time series. We additionally derive the direction of causal relationships (lead/lag relationships between variables at different time scales) and the strength of relationships (degrees of correlation). To implement the TY procedure, we have constructed vector autoregressive models (VAR Models) to deal with the integration or cointegration properties of the time series, and we used a modified version of Wald test statistics in the Granger causality test.

To summarize our contribution, we use the TY procedure for causality testing to understand the causal relationships between 27 financial markets, related to various world regions and countries, and the U.S. prediction markets contingent on the outcome of the 2008 U.S. elections. Our research provides insight into how financial markets, under various time granularities, react to the likely outcome of the U.S. presidential election. Finally, with our models we are able to comment on the efficiency of world financial markets, as defined by their ability to incorporate the likely outcome of U.S. presidential elections. Further, we are able to comment on the applicability of the Efficient Market Hypothesis (EMH) when it comes to U.S. presidential elections.

2.2 Theoretical Background

In this research, we use the Efficient Market Hypothesis, developed by the seminal study of Fama and Malkiel (1970), to help justify the impact of political information on financial markets. In the remainder of this section, we introduce three types of the EMH and then discuss the EMH testing in our research.

2.2.1 Three Types of EMH

An efficient financial market is one in which prices fully reflect all available information. Based on Fama's (1991) argument, a financial market is informationally efficient when the probability distribution of future prices, based on investors' relevant information, is equal to the true distribution of prices, considering all existing information in the market.

Consequently, the expected returns anticipated by investors are equal to true expected returns. Thus, EMH states that an asset's prices fully reflect all available information, and so it is impossible to beat the market which is not forecastable. In other words, stocks always trade at their fair value, which is based on incorporating all available information, including political information. The next section introduces three different forms of EMH, including the semi-strong form we use in this research.

Based on Jensen's (1978) definition, a market is efficient with respect to a specific information set, θ_t, if the investors cannot make more profit by using that information set. That is, as Malkiel (1992) says, the market is efficient with respect to a specific information set, θ_t, if revealing that information to all market players does not affect the price of shares. In these definitions underscore the importance of information sets, which can be used in financial market trading strategies.

Three forms of the EMH were identified in Fama and Malkiel's (1970) review. The first form, *Weak EMH*, posits that all price information is fully reflected in the asset prices. In other words, the information set, θ_t, contains only past and current asset prices. The second form, *Semi-strong EMH*, requires the inclusion of all available public information, in addition to the asset prices, in the information set. The last form, *Strong EMH*, asserts that the financial market reflects not only public information, but also private information that a limited number of investors may access.

The Strong EMH has been heavily criticized by a large number of scholars since it is very difficult to verify empirically. Despite the supposedly unbeatable market assumption of the Strong EMH, studies show that stock returns can be predicted by some macroeconomic factors such as interest rate (Balvers et al., 1990; Breen et al., 1989; Ferson and Harvey, 1993). In contrast, Fama and Malkiel (1970) shows that the Weak and Semi-strong forms of EMH are largely supported by empirical studies. Thus, the Semi-strong and Weak forms of EMH are preferred because they mitigate the difficulty of accessing and measuring private information. Additionally, there are two extensions to the EMH that investigates investor reaction when security prices do not adjust instantly: the Overreaction Hypothesis (OH), and the Uncertain Information Hypothesis (UIH) offered by DeBondt and Thaler (1985) and Brown et al. (1988), respectively.

2.2.2 EMH Testing

Despite the full acceptance of the EMH in the finance literature, many studies confirm the weak form of EMH in different local and regional financial markets, such as Latin America (Urrutia, 1995), Europe (Borges, 2010; Makovskỳ, 2014) and Asia (Kim and Shamsuddin, 2008; Cooray and Wickremasinghe, 2007; Aktas and Oncu, 2006). Furthermore, Gümüş and Zeren (2014) analyze the validity of the EMH in the G-20 countries, and show that the financial markets of the nine countries (the U.S., the U.K., Argentina, Australia, India, Italy, Japan, France and Germany) are efficient in the weak form.

Using the EMH approach, researchers investigate the impact of political event's information on the various financial markets. For instance, Tuck and Hon (2008), using the weak-form of EMH, show that the London Stock Exchange is an efficient market regardless of the political party forming the government. Moreover, Füss and Bechtel (2008) show the effect of 2002 German federal election on its stock returns.

In this research, we use the semi-strong EMH to assess the impact of U.S. presidential elections on global financial markets. The semi-strong form of the EMH may be difficult to test due to interdependence between information source and market (Cooray and Wickremasinghe, 2007); it is hard to determine the information source and if the information is instantaneously incorporated in the market. The volume of evidence supporting the weak form of the EMH suggests that even if markets as slightly inefficiency for a short amount of time, such inefficiencies are quickly corrected, allowing us to leverage the semi-strong version of the EMH in our study. The semi-strong form of EMH allows us to look at the likelihood of the U.S. presidential outcome, as measured by the prediction-market, and see how a country's financial market changes with this new information. Similarly, some country's financial markets may capture other information not currently incorporated in U.S. presidential election prediction market (say China purchases trillions of U.S. bonds), and this information may influence the predicted likelihood of the U.S. election outcome.

2.2.3 EMH Implications

One of the foremost implications of EMH is that stock traders are not able to beat the market because current stock prices reflect all currently available related information. However, historical data shows that the stock market can overreact to new information (De Bondt and Thaler, 1985; Howe, 1986). In other words, stock market inefficiency implies that prices will not react to new information immediately. Instead, stock markets may react gradually and potentially inadequately (i.e., underreact) or may overreact then adjust later.

Using the semi-strong form EMH, this research investigates whether publicly available information relevant to the U.S. presidential election is quickly incorporated in global financial markets and how markets react to this information. One of the issues that emerges from this investigation is that stock prices are not systematically over or under-valued, and traders should expect a normal rate of return (i.e., risk-adjusted return).

An implication of an efficient financial market is the possibility that there is no "best time" to purchase an asset. On the other hand, if financial markets are inefficient, asset price changes are not serially random, and traders can predict future prices based on observing the history of past prices. This research offers a method for identifying any inefficiency in global financial markets relevant to U.S. presidential election information that may be used to beat the market and make more profit than the market average.

2.3 Literature Review and Hypotheses

In this section we discuss related work to ours and formally state our hypotheses. Generally speaking, the outcome of presidential elections is not known prior to the election. Due to policy differences in the platforms of presidential candidates, the unknown election outcome is equivalent to unknown future policies. Given the connection between presidential election outcome and policy, we first discuss the literature on uncertain economic policy, then we turn our attention to presidential election outcome.

The link between politics and economy is heavily studied in the literature (Foerster and Schmitz, 1997; Alesina and Sachs, 1986; Lamb et al., 1997). Nordhaus (1975) develops the idea of the political business cycles to show the link between the government's policies and economic performance. Moreover, other studies show how macroeconomic outcomes such as unemployment, inflation are affected by a government's policies or political parties' promises (Carlsen, 2000; Hibbs, 1977*b*). The connection between political events and financial market performance is explored in many studies; in a survey paper Wisniewski (2016) introduces multiple studies that explore the links between politics and stock returns.

Among different political events, elections are potentially most important since economic policies are contingent on the outcome, leading to economic uncertainty prior to the election. Pantzalis et al. (2000) discuss the importance of the presidential election by arguing that elections empower voters to affect the course of economic policies and attract media attention that facilitates information dissemination into financial markets. Foerster and Schmitz (1997) and Foerster (1993) discuss the idea of the presidential election cycle, and they show that the U.S. stock market outperforms in the last two years of a presidential term. Moreover, the relationship of stock market movements and presidential elections outcome is studied by others (Chiu et al., 2005; Gemmill, 1992; Steeley, 2003; Brüggelambert, 2004). Additionally, there are several other studies which investigate the impact of other political uncertainties, such as strike, sanctions, terrorist attack, and legislation changes on financial markets (Bittlingmayer, 1998; Kim and Mei, 2001; Lin and Wang, 2004). A large and growing body of literature investigates the impact of the political environment on macroeconomic outcomes. For example, Hibbs (1977*b*) describes the role of partisanship policies on economic outputs, such as unemployment and inflation rates. Roubini and Sachs (1989) and De Haan and Sturm (1994) argue that a left-wing government has higher government spending. Similarly, Volkerink and De Haan (2001) show that right-wing administrations are more financially responsible. Our work fits in within the literature on the link between politics and the economy, and spans three streams of literature: 1) the impact of political elections on financial markets 2) political uncertainty and economic activity and 3) ex-post and ex-ante analysis of political events. We will discuss each of these three streams of literature in turn in our research.

We start by discussing the first stream of literature. Economic researchers attempted to evaluate the impact of a coalition government on economic outputs. They show that as coalition size grows (as measured by the number of political parties in the coalition), then so does public debt (Bawn and Rosenbluth, 2006; Persson et al., 2003; Alesina and Perotti, 1996).

Many financial market experts affirm that the U.S. stock market achieves better financial returns under Republicans than Democrats. For example, Niederhoffer et al. (1970) uses presidential election data from 1900 to 1968 and shows whenever the Republican won the elections, then the stock market performed better the day after the election. This finding, regarding the day-after market performance, is challenged in the literature by studies that investigate the relationship between political partisanship and financial markets in the short and long term.

The literature on short-term election impact considers the movement of market returns for a given investor's expectation of future government, which might be a right-wing or left-wing administration (Oehler et al., 2013; Roberts, 1990*a*; Białkowski et al., 2008;

Sturm, 2009; Homaifar et al., 1988). Additionally, Homaifar et al.'s (1988) investigates the defense-related industries' stock price movement in the period surrounding an election. The literature on long-term election impact mainly focuses on the relationship between political partisanship and stock markets in the long-term (Allvine and O'Neill, 1980; Huang, 1985; Foerster and Schmitz, 1997). These studies reinforce the idea of the political business cycle offered by Nordhaus (1975). Additionally, Stovall (1992) shows that the U.S. stock market returns are higher, specifically in the second half of the presidential terms, when the Republican party is in power. In contrast, Santa-Clara and Valkanov (2003) show that the Democrat's administration is better for the U.S. stock market returns, in which the average excess return from 1927 to 1998 under the Democrats and the Republicans are 2 and 11 percent, respectively.

Much of the previous research focused on the impact of political events on the U.S. stock market. However, there is a relatively large body of literature that investigates the other country's stock market. For example, Hudson et al. (1998) show that the U.K. stock market performs better under a right-wing government than a left-wing government; however, Döpke and Pierdzioch (2006) show that the German stock market is not sensitive to the political party in power; Cahan et al.'s (2005) argue that the New Zealand stock market's return was higher during right-wing administrations from 1931 to 2003 than left-wing administrations during the same period; and Wang and Lin (2008); Hung et al. (2007) empirically analyze that the congressional election has a negative effect on Taiwan stock returns. Further, Furió and Pardo (2012) support partisan politics theory since Spain's stock returns movement depends on the government's political orientation. Subsequently, Chiu et al. (2005) investigates the impact of political elections on foreign investor's trading behavior on the South Korea stock market.

Of the fist stream of literature that considers the impact of a domestic election on domestic financial market, none of them consider a prediction market as the estimate of the election. We revisit this idea when we discuss different lenses one may use when investigating political events (ex-ante vs. ex-post, see the end of this section). Moreover, as previously stated we consider the impact of U.S. elections on non-U.S. financial markets, i.e., domestic elections on non-domestic markets, something that to our knowledge is not explored previously.

We now move to the second stream of literature on political uncertainty and economic activity. Numerous studies find a negative effects of policy uncertainty on economic activity. For instance, Liu and Zhang (2015); Wang et al. (2014); Karnizova and Li (2014); Colombo (2013) discuss the financial aspects of economic policy uncertainty. Davis (2016) considers global economic policy uncertainty, a specific type of economic policy uncertainty, and develops the economic policy uncertainty (EPU) index based on newspaper coverage frequency. Furthermore, Baker et al. (2016) expands the EPU index to create a monthly index of Global Economic Policy Uncertainty (GEPU). Unlike our work, Davis (2016); Baker et al. (2016) use newspaper coverage to measure uncertainty, instead we use prediction market data.

Another manifestation of policy uncertainty is government policy uncertainty, for example associated with different presidential candidates. Pástor and Veronesi (2013) investigate the relationship between government policy, as written in news outlets, and stock prices. The authors conclude that increased government policy uncertainty leads to increased stock-market volatility.

As previously mentioned, presidential elections are a source of uncertainty. This suggests, that not only does uncertainty need to be quantified, in the form of risk, but also its impact on metrics of interest, for example financial markets, must be captured. Kelly et al. (2016) consider the impact of major political events on financial markets. The authors develop a model to isolate political uncertainty by exploiting its variation around major political events such as presidential elections. Kelly et al. conclude that major political events impact the equity options markets, and the effect spills over across countries, not just the country with the major political event. Our work differs from Kelly et al.'s (2016) study as the authors do not quantify the risk associated with the major political event, which we do using prediction market data. In addition, the authors use correlation analysis between political events and market data, instead we use time-series causality in our analysis. Considering the effect of a presidential election, the US election in our case, as a source of economic policy uncertainty, the remainder of this section discusses the economic analysis of elections and presents our hypotheses.

Our study relates to the stream of literature on political uncertainty and economic activity because presidential elections are a type of political uncertainty (which party will be in power). The literature in this stream proposes multiple measures for political uncertainty, but we are unable to find a single measure for presidential elections. In our study we use the instantaneous probabilities, as measured by prediction markets, as the measure for the political uncertainty associated with U.S. presidential elections.

We now turn our attention to the third stream of literature on ex-post and ex-ante analysis of political events. The relationship between economic variables and presidential election politics has been investigated in a significant number of studies. As a general overview, the literature divides into two main approaches, ex-post and ex-ante analysis of elections. The first group of studies, ex-post analysis, tries to find a connection between the election outcomes, the winner's political partisanship, and different macroeconomic variables (Fair, 1978; Fiorina, 1991; Bartels, 1988; Hibbs, 1987, 1977*a*; Erikson, 1989; Jacobson, 1990; Alesina and Roubini, 1992). In addition, Rogoff (1990); Alesina and Roubini (1992) studied political-business-cycle theories to understand how incumbent politicians try to manipulate the economy to increase their chance of re-election. Furthermore, Jacobson (1990); Alesina and Roubini (1992); Alesina and Rosenthal (1995); Hibbs et al. (1996) have shown the impact of partisan control of government on economic outcomes.

Other studies are made more accurate through using econometric models and time series analysis to highlight the relationship between election and particular economic sectors over a period. Unlike our work, in which we consider election predictions and not the realized outcome, all ex-post studies only consider the realized election outcome in their studies.

In addition to ex-post analysis, which is based on actual election results rather than forecasts, there are a few studies that take the ex-ante approach by using prediction-market data. Since the emergence of prediction markets, researchers have tried to use information aggregated by the prediction market in their economic analysis. Prediction markets are useful in drawing inferences about future events directly. However, many researchers have studied the relationships of prediction market data to other economic variables. In most cases, a time series of a particular contract in a prediction market is created and tested in an econometric model that has other economic variables. For instance, Roberts (1990*b*) studied the relationship between the daily rate of return for 58 defense industry securities and the probability that Ronald Reagan would win the 1980 presidential election. Prediction market data was obtained from an odds market managed by a famous odds maker, Ladbrokes, of that time. The author shows a significant relationship between returns for those defense firms and the probability of Ronald Reagan winning the election.

Herron et al. (1999) analyzed the link between the United States presidential election outcome, based on the 1992 Iowa Political Stock Market, and 15 different economic sectors, by selecting related Dow Jones Industry Group Indices to pinpoint those economic sectors that significantly correlated with the election outcome. Similarly, Knight (2004) used Iowa Political Stock Market prices to demonstrate how policy platforms capitalize on companies' stock prices by analyzing daily data of 70 firms during the six months prior to the 2000 U.S. election. Moreover, Snowberg et al. (2007) precisely estimate the response of equity, currency, and bond prices to changes in the majority party in the 2006 midterm U.S.

Congress elections.

Our research, an ex-ante study, is not based on the actual election results, but uses prediction market data as an accurate forecaster of election outcomes in its analysis. To our knowledge, this is the first study to investigate the causal relationships between U.S. presidential elections and global financial markets. Moreover, it is the first to use intraday time series in its analysis. This research differs from other ex-ante studies in that it uses prediction market data, as a time series in its causality analysis, instead of surveys and polls. The prediction markets are known to have more accurate forecasts than polls and surveys (Berg et al., 2008; Rothschild, 2009). In addition, prediction markets provide real-time continuous information about the election outcome, whereas polls and surveys provide only a snapshot fixed in time of an election's likely outcome.

2.3.1 Research Hypotheses

Roberts (1990*b*) articulates his research hypothesis about the expectation that future defense spending of 58 defense firms was contingent on the election of Jimmy Carter or Ronald Reagan in 1980. The results of the hypothesis tests show a direct relationship between the probability that Reagan wins the election and the expected increase in the returns of 58 defense firms.

Herron et al. (1999) hypothesized that, among 74 economic sector portfolios, no sector would be politically sensitive to the 1992 US presidential election outcome. However, when the authors tested the sensitivity of sectors thought to be politically sensitive, they found that there was, indeed, a correlation between the outcome of the election and some politically sensitive sectors. Knight (2004) determined the causal relationship between the outcome of the 2000 US elections, and the price of 70 securities that the author identifies as either Bush-favoring or Gore-favoring. Also, Snowberg et al. (2007) explored financial market responses to the outcome of the 2006 midterm election by investigating the extent to which partisan majorities in Congress influenced economic policy.

Most of the literature looks at the relationship between U.S. election outcomes and domestic financial markets. In our work, however, we will present both the existence and direction of relations between the outcome of U.S. elections on global financial markets. We use prediction market data to identify the real-time probability estimate on the outcome of the elections. Similarly, we use a market index to determine the real-time value of an entire regional or country specific financial market.

In this research, each hypothesis had been constructed for the different set variables related to a particular prediction market and global financial market time series. For example, if we choose " Democratic Party Candidate to Win 2008 Presidential Election" (DEM2008) as a prediction market variable and choose "Dow Jones Canada Index" (CADAWD) as a financial market variable, then the general form of the null and alternative hypothesis of the Granger causality test are as follows:

$$Lead\ Hypotheses = \begin{cases} H_0: & \text{DEM2008 does not Granger cause CADOWD} \\ H_1: & \text{DEM2008 does Granger cause CADOWD} \end{cases} \tag{2.1}$$

$$Follow\ Hypotheses = \begin{cases} H_0: & \text{CADOWD does not Granger cause DEM2008} \\ H_1: & \text{CADOWD does Granger cause DEM2008} \end{cases} \tag{2.2}$$

The Lead hypothesis investigates the existence of relationships between the prediction market and financial market, in which a prediction market leads a financial market. In contrast, the Follow hypothesis investigates those situations in which a prediction market follows a financial market.

Lead and Follow hypotheses have been tested for 594 pairs, generated by the combination of two prediction markets, 27 financial markets in 11 different granularities. Two iterative testing strategies, Rolling Day and Moving Window (see Section 2.5.2), were used to test the hypotheses over $790,000$ times. The Rolling Day strategy tested over 430 iterations per pair; and the Moving Window strategy tested 1030 times per pair. We aggregated the results of the $790,000$ tests in the following mutually exclusive causal relationships[1]

- Lead Effect (PM $\rightarrow$ SM): a unilateral causal relationship exists by which changes in prediction markets lead to changes in financial markets; consequently, prices in prediction markets affect prices in financial markets.
- Follow Effect (SM $\rightarrow$ PM): a unilateral causal relationship exists by which changes in financial markets lead to changes in prediction markets; consequently, prices in prediction markets follow prices in financial markets.
- Feedback Effect (PM $\leftrightarrow$ SM): a bilateral causal relationship exists between financial markets and prediction markets, so both markets affect each other at the same time.
- Neutral Effect (PM $\otimes$ SM): a neutral causal relationship exists between financial markets and prediction markets, which means that the two types of markets are independent of each other and there is no causal relationship.

2.4 Data Description

We use different data sets relevant to this research: prediction market data and global financial data. Prediction market data show electoral probabilities from the United States 2008 presidential election campaign, for both the Democratic and Republican parties led by Barack Obama and John McCain, respectively. Global financial market data were selected from 27 different countries or regions' Dow Jones Global Indices during the same election campaign period (the list of all 27 indices is presented at the end of Section 2.4.2).

[1]PM stands for Prediction Market and SM stands for Stock Market/Financial Market.

2.4.1 Prediction Market Data

The first set of data was acquired from the intrade.com website. Intrade, founded in 1993, was later acquired by Tradesports in 2003. The Intrade website was eventually shut down in March 2013. Intrade, at one time, was one of the leading prediction market platforms, and all its trade data is publicly available. Intrade was a prediction-market trading exchange that had different contracts for many political events. Among the Intrade electoral contracts, we choose DEM2008 and REP2008 (Table 2.1), which are interpreted as the probabilities assessed by all market participants of a Democratic party victory and a Republican party victory in the 2008 US presidential elections, respectively.

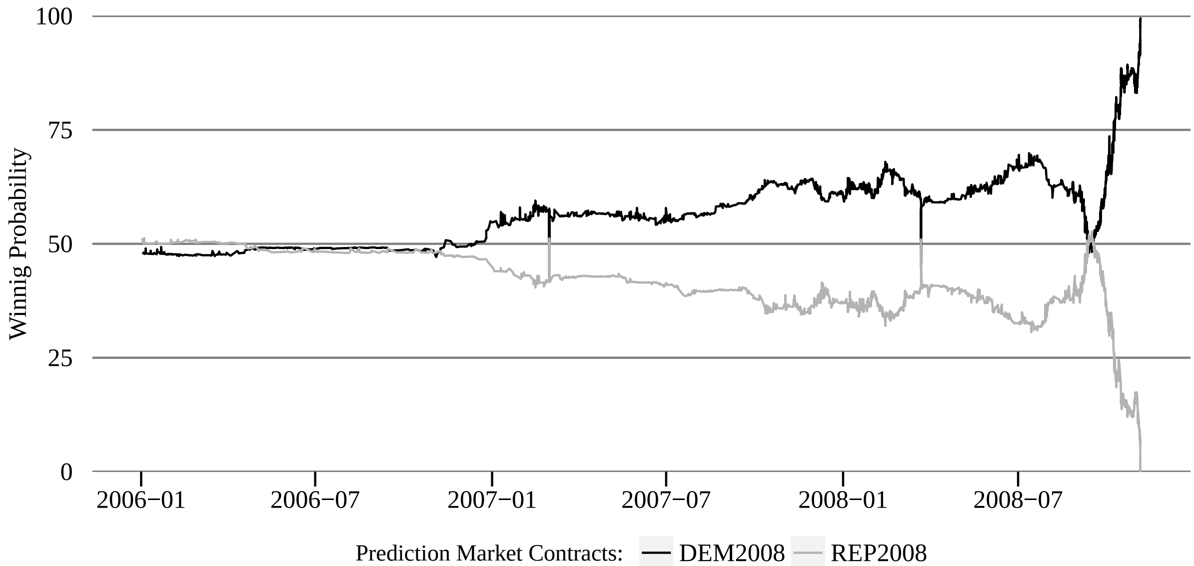
Table 2.1: Prediction market contracts description

Contract name	Contract description
DEM2008	Democratic party winning 2008 the U.S. presidential election
REP2008	Republican party winning 2008 the U.S. presidential election

Each prediction market time series (DEM2008 and REP2008) starts at the beginning of 2006 and ends on November 4, 2008, which was the United States presidential election date. These two time series are probability of winning Figure 2.1 shows the movement of the two parties' winning probabilities (DEM2008 with a black line and REP2008 with a grey line).

Intrade contracts represent a "winner-takes-all" type (a winner is paid 100 cents for each winning contract held), in which a contract's price represents the expected probability of an event outcome. Regularly, the probability of winning the election for each party sums to unity at any point in the time series. But there are some cases in which these two probabilities do not complement each other due to website friction, a percentage of all winning values are kept by the Intrade. As Figure 2.1 shows, both parties started with a

Figure 2.1: Intrade Prediction Market Data - 2008 U.S. election



winning probability of 50%, and traders aggregated all information into the share's prices during the campaigning. By Intrade prediction market's design, players were free to trade contract shares any time to make sure that the trader's information was aggregated as soon as possible, and the winner took all the value of its contracts. Eventually, the price of the Democratic party approached 100, and the price of the Republican party approached 0 on election day.

2.4.2 Global Financial Market Data

The second category of data in this research consists of the minute-by-minute indices for 27 countries, available in the Dow Jones Global Indices (DJGIs) family. One important feature of DJGIs is the comprehensive coverage of each country across different economic sectors. Therefore, we assume that DJGIs are the proxy of the overall financial market's situation.

DJGIs indices are weighted by float-adjusted market capitalization, in which the indices calculation utilize the divisor methodology used in all S&P Dow Jones Indices' equity indices (Jones, 2019). The formula to calculate capitalization-weighted indices, such as all DJGIs indices, is as follows:

$$Index\ Level = I = \frac{\sum_i P_i * Q_i}{Divisor} \tag{2.3}$$

where P_i is price of each stock and Q_i is the number of shares used in the index calculation. The denominator is the divisor, for example when the sum of the numerator is US$25 trillion and the divisor is US$ 10 billion, the index level will be 2500. Moreover, In the float-adjusted weighted methodology, a change in the index level is defined by a LasPeyres index as follows:

$$Index\ Level\ Change = \frac{I + \Delta I}{I} = \frac{\sum_i P_{i,1} * Q_{i,0}}{\sum_i P_{i,0} * Q_{i,0}} \tag{2.4}$$

where I is the index level; P_i and Q_i are price and the float adjusted share count of asset i. Based on equations (2.4) and (2.3), total returns indices are defined as follows:

$$\frac{I_{TR,t}}{I_{TR,t-1}} = \frac{I_{t-1} + \Delta I_t + \sum_{i,t}(D_{i,t} * Q_{i,t})}{I_{TR,t-1}} \tag{2.5}$$

where I_{TR} is the total return index level and $D_{i,t}$ is the dividend for asset i on dividend execution-date (reinvestment date) t. In particular, to figure the rate of return, we need to know the starting price and ending price for the index for a specific period of time.

In this research, we use DJGIs level data in our time-series analysis as descried above. However, we take first difference of all index time series to correct the non-stationary issues discussed on the Section 2.5. As such we are looking at price changes in each index and comparing these to changes in the probability of the presidential election outcome.

All selected DJGIs indices are calculated in U.S. dollar currency and contain trading days from January 1, 2006, to November 4, 2008. We have kept non-trading days as we have them in the prediction market time series as well. The original intraday historical data, acquired from www.kibot.com, is in five-minute granularity, and we created other time series with different time granularities based on the original time series.

The selected Dow Jones indices time series are as follows: Americas (A1DOW), Latin America (A3DOW), Latin America ex-Mexico (A3DOW), Austria (ATDOWD), Australia (AUDOWD), Belgium (BEDOWD), Brazil (BRDOWD), Canada (CADOWD), Germany (DEDOWD), Denmark (DKDOWD), Europe (E1DOW), Europe ex-U.K. (E2DOW), Europe - Nordic (E3DOW), Finland (FIDOWD), France (FRDOWD), Greece (GRDOWD), Hong Kong (HKDOWD), Indonesia (IDDOWD), Ireland (IEDOWD), Italy (ITDOWD), Japan (JPDOWD), Mexico (MXDOWD), Malaysia (MYDOWD), Netherlands (NLDOWD), Norway (NODOWD), New Zealand (NZDOWD), China (DJCHINA).[2]

2.4.3 Time Series Pairs, Granularity and Transformation

To investigate the relationship of the global financial markets and the U.S. presidential prediction markets, for each pair of time series, we matched one of the prediction market time series (DEM2008 and REP2008) with the closing price, the end of the time granularity considered, of the selected DJGI indices. The process of creating time series is as follows:

1. Pairs: considering two prediction markets series and 27 financial indices, we created 54 pairs of time series. For example, the CADOWD-DEM2008 series (Figure 2.2) is the combination of the Dow Jones Canada Index stock market time series and the prediction market time series for the Democratic Party Candidate to Win 2008

[2] For more information on Dow Jones Global Indices® and download a comprehensive list of all DJGI indices, visit the following website: https://ca.spindices.com/documents/index-policies/dj_vendor_codes.xls?force_download=true

Presidential Election. The new time series has a time stamp, corresponding to the selected time granularity, and the closing prices for both the stock market and prediction market time series.

2. Granularity: we aggregated pairs of time series to 11 different frequencies in the standard open-high-low-close (OHLC) format. We generated time series of 5, 10, 15 and 30-minute intervals, and time series of 1, 2, 4, 8, 12 and 16-hour intervals. Finally, we created a daily time series for each pair. In total we used 594 pairs of time series in our empirical analysis based on closing prices for each time interval considered.

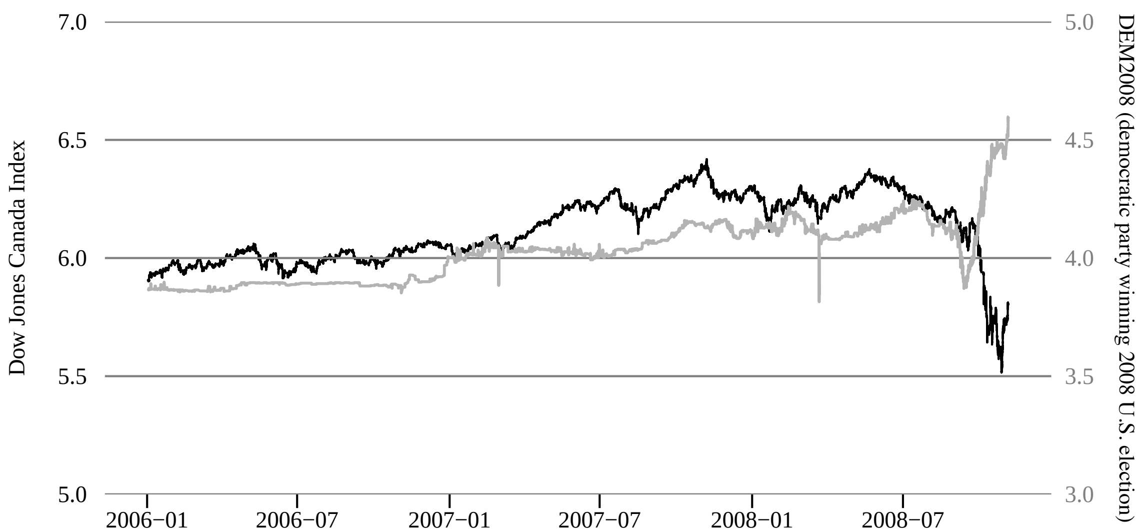
3. Filling gaps: pairing prediction-market and financial-market time series of various granularity may generate some gaps in both time series. For example, in a weekday, we have financial market price data for every 5 minutes, but we might have no data for the same times in the prediction market. In such a case, we fill the time series gaps with the closest previous price.

4. Transformation: we use the natural logarithmic transformation of all time series to make elasticity calculation easier, by allowing us to interpret small changes in the transformed variables as percentage changes. In addition, such logarithmic transformation can help to stabilize the variance of the time series as shown by Lütkepohl (2005).

2.5 Statistical Methods

This section highlights the statistical methods used to investigate the relationship between the U.S. election prediction markets and the selected global financial markets.

Figure 2.2: Pair of CADOWD-DEM2008 in five minutes interval



Generally, the basic approach currently used in research on causality between variables is the Granger causality test, which is based on Granger's (1969) original work.

Granger causality testing is heavily used in the literature to understand the existence of causal relationships, to infer causality directions, and to identify lead and lag variables. Many other researchers have expanded Granger's idea and offered some improvements, including Sims (1972), Toda and Yamamoto (1995), Dufour and Renault (1998), Lütkepohl (2005). But, conventional Granger causality test has some limitations as follows:

- Granger causality requires time-series to be stationary. However, it is well known financial time series are not stationary (Maddala and Lahiri, 2001) and as such conventional Granger causality testing suffers from problems such as spurious regression results (Gujarati and Porter, 2009). Also, Gujarati and Porter argue that the F-test is not valid when the variables are cointegrated.

- Granger causality test is sensitive to model specification and the number of lags. Granger testing would reveal different results if a causal factor was relevant and was not included in the model (Seth, 2007; Gujarati and Porter, 2003).

- Hu et al. (2012) discuss the Granger causality's shortcomings in time domain (i.e., Auto-regressive models) and show how Granger causality testing fails to determine the strength of causality when there is bi-directional causality between two time series.

To remedy the non-stationary problem, the Error-Correction Model (ECM) and the Vector Autoregression Error-Correction Model (VECM) are respectively offered by Engle and Granger (1987) and Johansen and Juselius (1990) as alternatives to conventional Granger causality testing. However, these alternative tests are highly sensitive to the value of parameters in finite samples and also to the results of some pretests, such as the unit roots and cointagrating ranking necessary in these tests. Distortions may, therefore, occur associated with non-stationary and cointegration pretests. Due to these distortions, we avoid using ECM and VECM tests as results from these tests may not be reliable due to the lack of usual asymptotic distribution in integrated or cointegrated series (Toda and Yamamoto, 1995; Zapata and Rambaldi, 1997; Rambaldi and Doran, 1996).

All things considered in this study, we use the Toda-Yamamoto causality test, which is an adjusted Granger non-causality test, proposed by Toda and Yamamoto (1995) to address non-stationary and co-integration problems of financial time series. Also, we consider a Zivot-Andrew test to detect any structural break. In the next section, we briefly discuss TY procedure.

Granger causality testing assumes linear interactions between a model's variables, and this linearity assumption might lead to erroneous conclusions when nonlinear interactions occur. Alternatively, non-linear Granger causality testing methods have been proposed by Hiemstra and Jones (1994) and extended by Diks and Panchenko (2006). However, these methods cannot be applied to non-stationary data Hassani et al. (2010). Moreover, transforming a dataset into a stationary process (e.g., the first difference of a dataset) may remove information about the long-run relationship between variablesStern and Enflo (2013). As such, as we consider nonlinear time series, we only consider linear relationships between variables.

2.5.1 Toda-Yammamoto Causality Test Procedure

The limitations of the the Granger causality test due to non-stationary time series could lead to spurious regression problems. Furthermore, when the variables are integrated or cointegrated, the common F-test process is not valid since the test statistics do not have a standard distribution.

The Toda-Yamamoto test came about in order to fix the problems of invalid asymptotic critical values when we use causality tests with non-stationary or cointegrated time series.

Toda and Yamamoto (1995) proposed a simple procedure requiring the estimation of a VAR model, which guarantees the asymptotic distribution of the Wald statistic that has asymptotic χ^2 distribution. The TY testing procedure can deal with the integration or cointegration properties of the process. The Toda-Yamomato Granger causality test is an augmented VAR model for two time series, x and y, as follows:

$$x_t = a + \sum_{i=1}^{k} \alpha_{1i} x_{t-i} + \sum_{i=k+1}^{k+d_{max}} \alpha_{2i} x_{t-i} + \sum_{i=1}^{k} \beta_{1i} y_{t-i} + \sum_{i=k+1}^{k+d_{max}} \beta_{2i} y_{t-i} + \varepsilon \tag{2.6}$$

$$y_t = a + \sum_{i=1}^{k} \gamma_{1i} y_{t-i} + \sum_{i=k+1}^{k+d_{max}} \gamma_{2i} y_{t-i} + \sum_{i=1}^{k} \delta_{1i} x_{t-i} + \sum_{i=k+1}^{k+d_{max}} \delta_{2i} x_{t-i} + \varepsilon \tag{2.7}$$

The TY procedure suggests creating a lag augmented VAR(k+d_{max}) by using a modified version of Wald test statistics (MWald) that has an asymptotic χ^2 distribution. In this case, d_{max} is the maximal order of integration and k is the optimal lag length of the system.

In this approach, the VAR model with the correct order of k is augmented by the maximum order of integration, d_{max}, and we ignore the coefficients of last d_{max} lags. So, "y does not Granger cause x" implies that in equation (2.6) $\beta_{11} = \beta_{12} = ... = \beta_{1k} = 0$; similarly, in equation (2.7), "x does not Granger cause y" if $\delta_{11} = \delta_{12} = ... = \delta_{1k} = 0$.

As Figure 2.3 shows, to implement the TY procedure, we follow three main steps:

1. Identifying the maximum order of integration, d_{max};
2. Determining the optimal lag length, k;
3. Testing for Granger causality, using modified Wald test for VAR($k + d_{max}$).

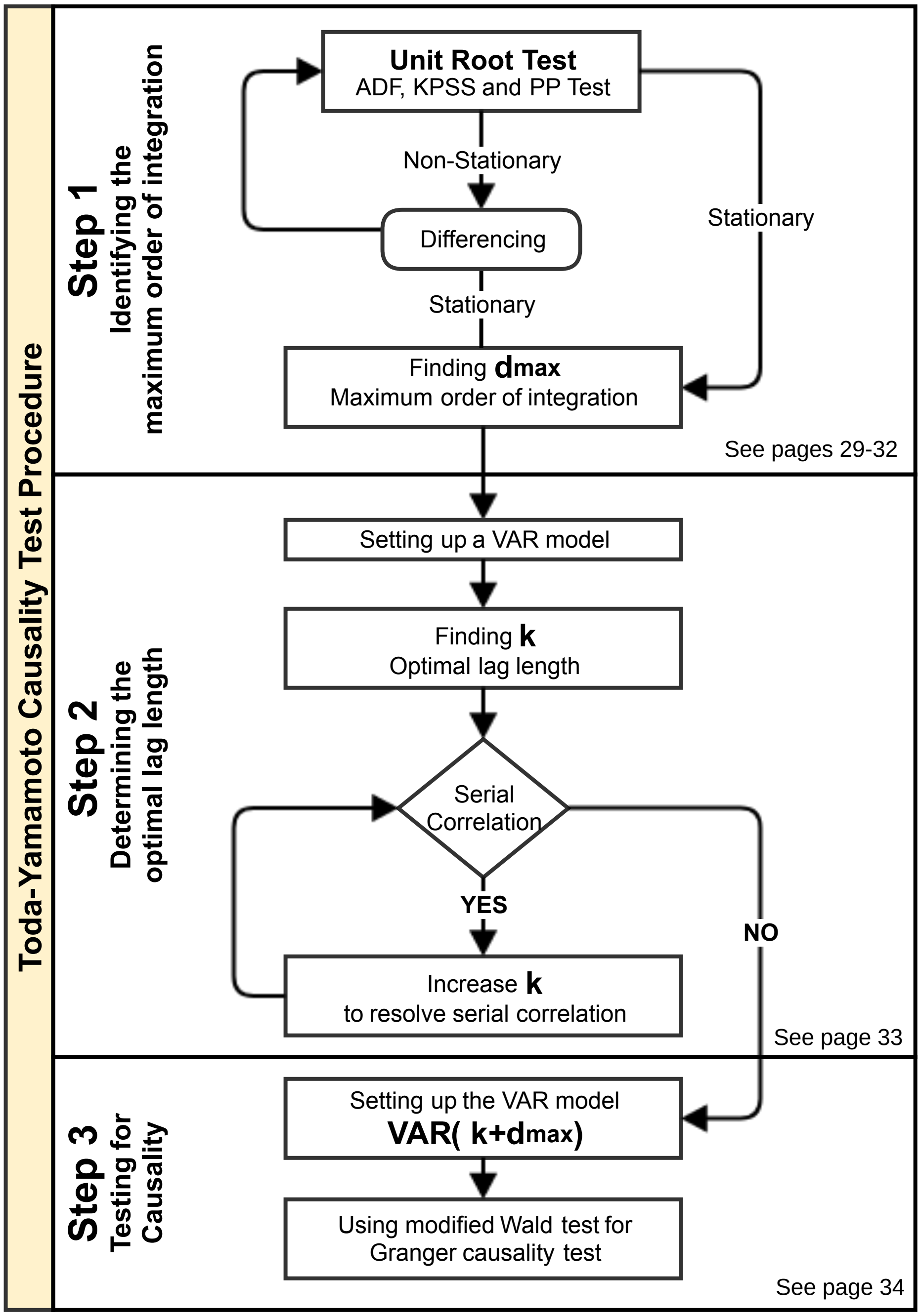
Step 1- Identifying the Maximum Order of Integration

Order of integration, $I(d)$, is the minimum number of differences required to obtain a stationary time series. To determine whether a time series is stationary or not, we use a unit root test. If the unit root test result shows a degree of integration in the selected time series, then we difference the time series, meaning that we compute the differences between consecutive observations and retry the unit root test. We continue differencing until we produce a non-stationary time series. Finally, the number of times we must make a difference is the order of integration.

Finding the order of integration must be done for series available in the VAR model, and we select the maximum order as the VAR's maximum order of integration. For example, if we have $I(1)$ for a prediction market series and $I(2)$ for a financial market series, then the VAR's maximum order of integration will be $I(2)$.

To investigate unit root in our time series we use three different tests: The augmented Dickey-Fuller (ADF) test offered by Dickey and Fuller (1979), the Phillips-Perron (PP) test

Figure 2.3: Toda-Yamamoto Causality Test Procedure



offered by Phillips and Perron (1988), and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test offered by Kwiatkowski et al. (1992).

Selecting the true lag length, p, in an ADF test is important. Too small a lag length will cause serial correlation in the error terms and thereby bias the test results. On the other hand, choosing a large lag length will reduce the power of the test (Zivot and Wang, 2006). Moreover, according to Gujarati and Porter's (2003) argument, selecting too many lags in the model will decrease the degrees of freedom and cause over-fitting and multicollinearity (correlation of the independent variables) problems, whereas choosing too few lags will lead to specification errors (correlation of the independent variables with the error term). The common strategy for handling this issue is using the information criteria offered by many scholars, such as AIC (Akaike, 1973), BIC (Schwarz, 1978), and HQ (Hannan and Quinn, 1979).

Based on Gonzalo and Pitarakis (2002), penalty term in the AIC leads to a non-zero limiting probability of over fitting when $T \to \infty$, where T represents the total number of observations. Furthermore, Davidson and MacKinnon (2004) found that the AIC criterion may fail to choose the most parsimonious model (i.e., the model with the fewest parameters to estimate). BIC, however, might lead to an overly parsimonious situation by under fitting the model. Additionally, Gonzalo and Pitarakis (2002) show that BIC and HQ lead to consistent estimates and are better than AIC in models with a large number of observations.

Considering the controversial problems of over fitting and under fitting a model by using the traditional IC, Ng and Perron (2001) proposed a modified information criterion (mIC) to consider a good size and power of unit root test in selecting true lag length. Therefore, in this research, we use mAIC information criterion to choose an appropriate lag length in the ADF unit root test as follows:

$$mAIC(p) = ln(\hat{\sigma}_p^2) + \frac{2(\tau_T(p) + p)}{T - p_{max}} \tag{2.8}$$

Where:

$$\tau_T(p) = \frac{\hat{\pi}^2 \sum_{t=p_{max}+1}^{T} y_{t-1}}{\hat{\sigma}_p^2} \quad and \quad \hat{\sigma}_p^2 = \frac{1}{T - p_{max}} \sum_{t=p_{max}+1}^{T} \hat{\varepsilon}_t^2 \tag{2.9}$$

To calculate mAIC in equation (2.8), we need to determine p_{max},which is the maximum lag length needed to calculate ICs for finding the appropriate lag length. Schwert (2002) suggests a rule of thumb, which is investigated with a Monte Carlo study, to determine p_{max} as follows:

$$p_{max} = 12 \times \left(\frac{T}{100}\right)^{\frac{1}{4}} \tag{2.10}$$

There is a three-step process for finding p : 1) calculate p_{max} , 2) calculate mAICs for lag length from 1 to p_{max}, and 3) select the best p where mAIC is the minimum number.

The other two unit root tests, Philips-Peron (PP) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS), are used to cross check the result of ADF testing. Hobijn et al. (2004) argue that ADF unit root testing has low power against stationary and highly auto-regressive series. In other words, we may not find the stationarity of a series with ADF testing. To deal with highly auto-regressive models, Hobijn et al. argued the good performance of the KPSS test based on a Monte Carlo investigation. Moreover, ADF testing assumes that u_t is a stationary white noise process and addresses the serial correlation of error terms by adding some lagged variables into the model. Whereas we have some challenges to determine true lag length, Mahadeva and Robinson (2004) argue some advantages of PP testing over ADF testing because the former's non-parametric approach does not need to select any lag length to address serial correlation problems in the error term.

Therefore, in this research, we conclude the non-stationary property of our time series when the results of all three tests (ADF, PP and KPSS) unanimously confirm the non-stationarity of the selected time series. If they do not agree, then we flag that test and do not continue the TY process.

Step 2- Determining the Optimal Lag Length

After determining d_{max} in the first step of the TY process, we need to find the optimal lag length of the VAR model, which includes the following tasks:

1. Selecting the lag length, p, based on the information criterion.

2. Checking for the possible serial correlation of the VAR(p) model.

First task: VAR lag order selection

In the first task for VAR model lag-order selection, we use the BIC and HQ information criteria based on Scott Hacker and Hatemi-J's (2008) simulation study that finds BIC and HQ have the best performance in many cases. First, we determine the maximum lag length, p_{max}, by the rule of thumb presented in equation (2.10). Second, we calculate HQs and BICs for VAR(1) to VAR(p_{max}). The minimum number of calculated BICs and HQs will help us to understand the offered lag orders, p_{BIC} and p_{HQ}, based on both information criteria. If $p_{BIC} = p_{HQ}$, then we can determine the VAR model lag length, p, as $p = p_{BIC} = p_{HQ}$. Otherwise, if $p_{BIC} \neq p_{HQ}$ we use a Likelihood Ratio (LR) test to choose the VAR model lag length, p, from p_{BIC} or p_{HQ}. This LR test is recommended by Hatemi-J and S. Hacker (2009) when ICs suggest different lag orders for a VAR model.

Second task: Serial correlation

In the second task, we have to make sure that the residual of the VAR(p) is not correlated with independent variables. If we find serial correlation in our VAR(p) model, we need to increase the order of the VAR model to $p + 1$ and test it for serial correlation.

This process must repeated until the serial correlation problem is resolved in the VAR model.

This research uses the Bruech-Godfrey (BG) test, which is offered by Breusch (1978); Godfrey (1978), to test serial correlation of the VAR models. We use Breusch-Godfrey test to check the possible serial correlation of VAR models for up to 10 lags. The choice of 10 lags for an autocorrelation test is based on Hyndman and Athanasopoulos's (2014) suggestion when dealing with non-seasonal data.

Step 3- Testing for Granger Causality

In conventional Granger causality tests, zero restrictions on the coefficients can be checked by the Wald test, which has asymptotic χ^2 distribution, a standard tool for VAR models, when we have stationary data. However, according to Toda and Yamamoto (1995), if the time series are integrated or cointegrated, then the standard asymptotic theory is not applicable and the standard Wald test-statistic does not have asymptotic χ^2 distribution in the level VAR (the original VAR model with no differencing). We thus create VAR(p+d_{max}) and use Wald test for the first p variables without considering d_{max} lags. This method guarantees the asymptotic distribution of the Wald statistic, which is asymptotically χ^2 distributed with p degrees of freedom under the null hypothesis.

2.5.2 TY Testing Strategies

We implemented the TY procedure with two different iterative strategies, Rolling Day and Moving Window. Using these two strategies helps us investigate how financial markets incorporate the U.S. election outcome in different time frames (i.e., long term and short term). Additionally, using these iterative strategies helps us to find and exclude structural breaks that might exist in our time series. For example, under the Moving Window strategy, we subset the time series based on windows width and implement the TY procedure only on data in that window to find any changes in causality relationships in the short term. The following two sections explain these two strategies in greater detail.

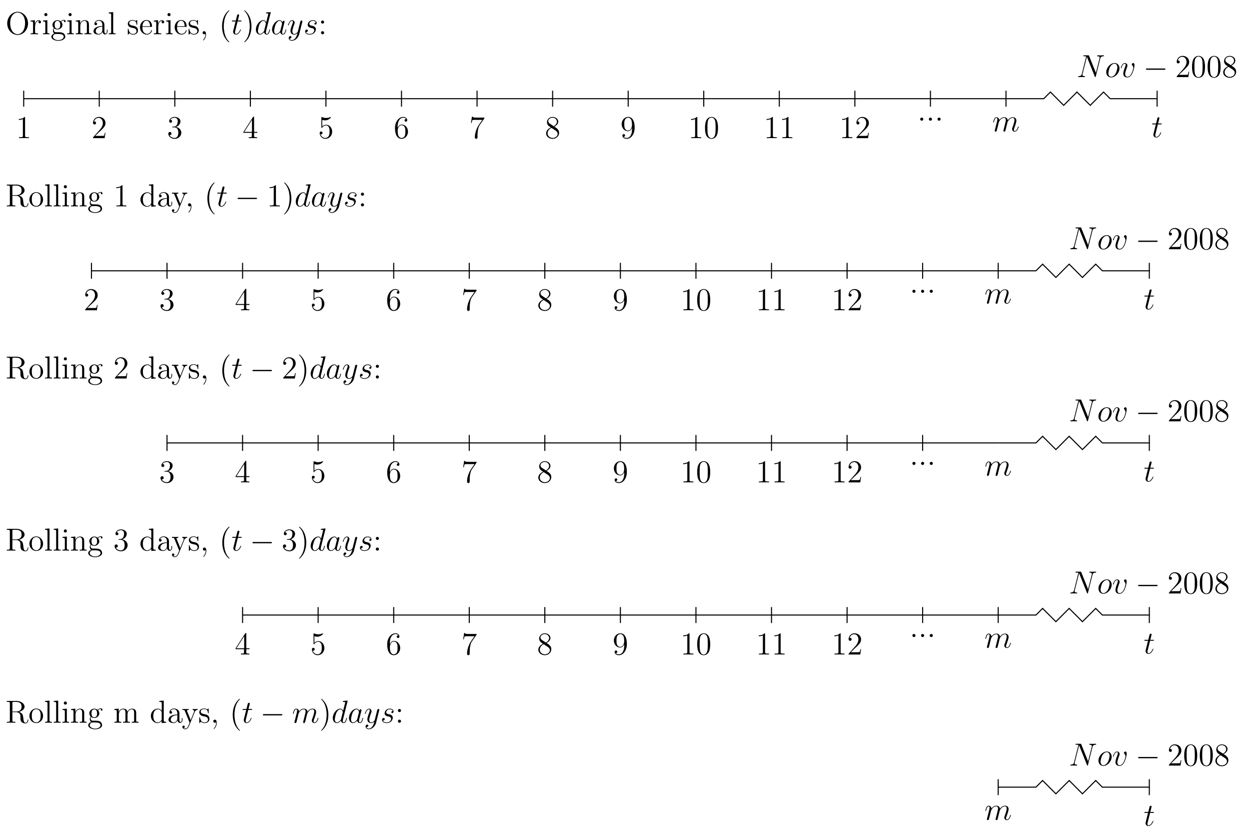
Rolling Day Strategy

With the first strategy, Rolling Day, in each iteration we exclude observations related to the first day of the original time series and create a new time series. Then we test the TY procedure to determine Granger causality in the new time series. We continue each iteration by excluding the first-day observations until, in the last iteration, we have only observations related to the last few days of our original time series.

We use the rolling horizon due to the variability and irregular structure of prediction market time series. The number of trades per day at the beginning of both prediction market time series (i.e., 14 months before the election date) are lower than that at the end because trading increases remarkably near the election date. Simultaneously, financial market time series are regular during market hours, with five-minute granularity. In other words, when we regularize a pair of time series, the limited number of changes at the beginning of prediction market time series may affect lag order selection and causality results. Moreover, Hacker and Hatemi-J (2006) show that the modified Wald (MWALD) test is sensitive to and performs poorly with small sample sizes. Thus, by excluding observations related to a specific day at the time-series beginning in each iteration, we remove those low-change areas, and by moving forward to the next iterations, we consider the different structures of time series in our analysis to reduce the risk of lag-order-selection errors.

As Figure 2.4 illustrates, the original series includes t days, and the total number of observations is T. In the first iteration, the first subset has $t - 1$ days, and the number of observations is $T - n_1$, where n_1 is the total number of observations in the first day, which we excluded from the original series. In the last iteration, the last subset has $t - m$ days, and the number of observations is $T - n_m$, where n_m is the total number of observations from the first day to the m^{th} day of the selected time series. Here, m is set to $t - 1$, $t - 5$ and $t - 30$ based on time series granularity to avoid a low number of observations in the last subset.

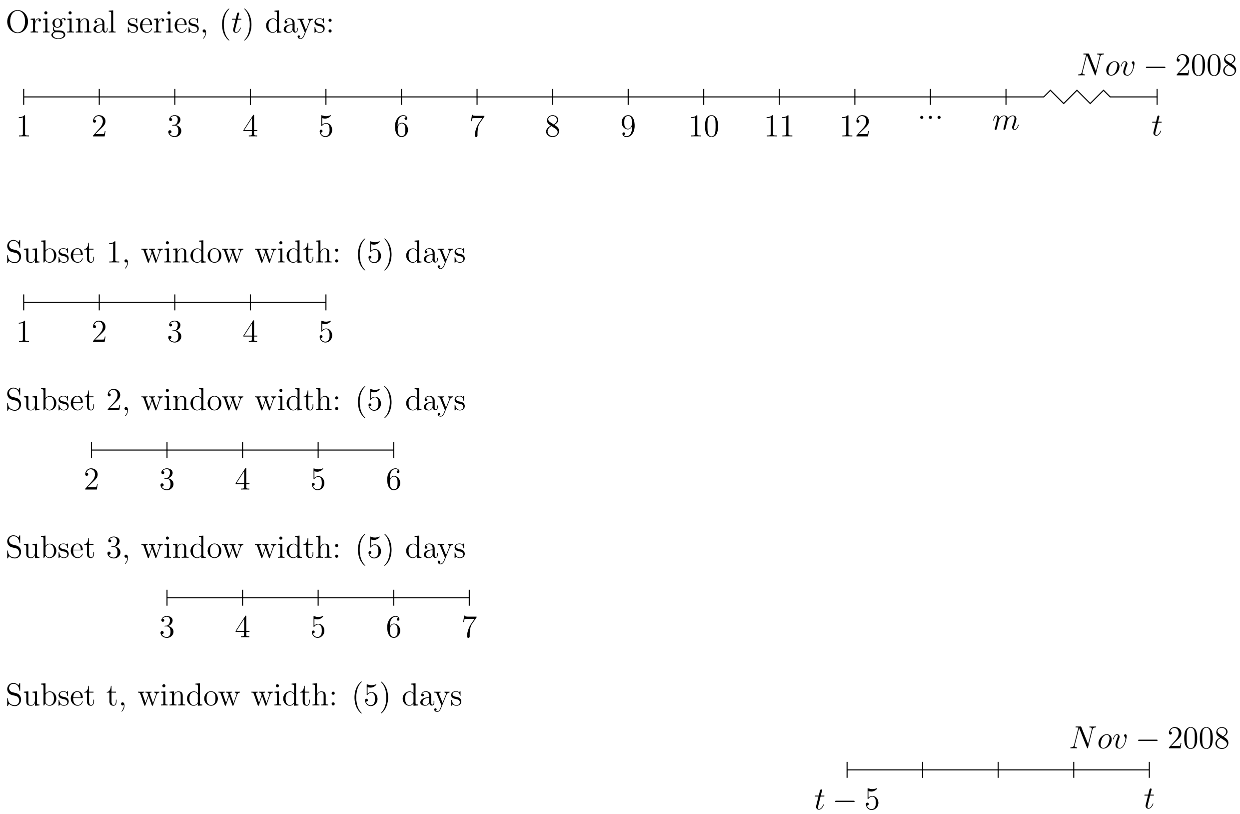
Figure 2.4: Rolling Day subset selection strategy



Moving Window Strategy

With Moving Window, the second strategy to implement the TY procedure, we set the window widths to 3, 5, 10, 15, 30 or 60 days. Then, for the first iteration, we start from the first day of the original time series and select those observations that are in the scope of the selected window. In each iteration, the TY process is tested on the selected observations. For the next iteration, we move our window forward by one day and test TY on the new window's observations. We continue to move our window forward to the end of the time series and we do the TY process on each iteration.

Figure 2.5: Moving Window subset selection strategy



As Figure 2.5 shows, the original series includes t days, and the total number of observations is T. Assume that we set the length of each moving window to 5; the first subset in the first iteration has 5 days and the number of observations is $n_{5,1}$, where $n_{5,1}$ is the total number of observations in the first 5 days of the original series, and the next iteration has $n_{5,2}$ observations. In the last iteration, the number of observations is $n_{5,t-5}$, which is the total number of observations from the last five days of the selected time series.

2.6 Causality analysis on the Canadian financial market, case study

In this section, we show the results of TY procedure with two testing strategies in detail. With each strategy, we provide TY procedure results for the CADOWD-DEM2008 pair as a sample and then summarize the test results for the other pair, CADOWD-REP2008.

2.6.1 Rolling Day Strategy Results

As an example, a pair, CADOWD and DEM2008, with 5 minute granularity has 475 days and 110125 observations; m is set to $t-5$, and the number of subsets is 470. Thus, we perform the TY process 470 times for the CADOWD-DEM2008 pair with the Rolling Day strategy. In this case, we choose CADOWD-D-M5-1 as a unique ID of the first iteration, in which D stands for DEM2008, $M5$ stands for 5-minute granularity, and the numeral (1) shows the iteration ID.

Based on the TY procedure presented in Figure 2.3, to run the Granger causality test for CADOWD-D-M5-1, we start with the unit root test as the first step. In this case, the number of observations is $74,529$, and the maximum lag for the ADF test is 62, calculated based on the equation (2.10). For the ADF test, we need to use a proper lag order, which is selected using the steps outlined in Section 2.5.1. In this research, for the ADF test, we use a lag order offered by $mAIC$ criterion which is equal to 1 for the CADOWD time series and 12 for the DEM2008 time series.

The null hypotheses of the ADF and PP say that there is a unit root present in the time series; on the other hand, the null hypothesis of KPSS says there is no unit root in the time series. According to the p-values, we cannot reject the null hypothesis of ADF and PP tests. In addition, the p-value of the KPSS test confirms that we can reject the null hypothesis. Thus, all three tests unanimously confirm the existence of a unit root in both series.

Based on our TY procedures, if we find a unit root in the time series, we need to transform the original time series, $I(0)$, by making a first difference to create a new time series, $I(1)$. Then, the results of ADF, PP and KPSS tests are consistent with one another indicating that both transformed time series are stationary.

By resolving the unit root problem of both time series, CADOWD and DEM2008, we conclude that the maximum order of integration, d_{max}, is equal to one for the pair of CADOWD-DEM2008-M5-1.

After finding the value of d_{max}, we need to the find optimal lag length, k, of the VAR model with the selected pair. First, we need to determine the model's lag length based on its information criteria. We set up a VAR model for the CADOWD-D-M5-1 series with the maximum lag of 62, calculated based on the equation (2.10), and we calculate the information criteria, $HQ(n)$ and $SC(n)$ which are 14 and 7, respectively.

Since the reported $HQ(n)$ and $SC(n)$ are not equal, so based on the discussion in Section 2.5.1, we use the Likelihood Ratio test (LR Test) to decide which information criteria are appropriate for selecting the VAR model lag length. For the LR test, first, we set up two VAR models, VAR (lag=14) and VAR (lag=7), and then we calculate the log-likelihood of both, which are 736723.5 ($df = 32$) and 736772.3 ($df = 60$), respectively. In this case, when $HQ(n) < SC(n)$, the lag length of the restricted model is set to 7 and the unrestricted model to 14, and the degree of freedom for the LR test is 60. Finally, the test statistic of the LR test is 97.63. If we then compare the LR test statistic with the chi-square critical value of 77.93, calculated with a 95% significance level and 59 degrees of freedom, we conclude that the VAR lag length is equal to 14, since the LR test statistic is less than the critical value. Thus, we need to take the unrestricted model lag, 14, as the appropriate lag length for the VAR model.

We set up a new VAR model with the selected lag length of 14. To ensure that the model has no serial correlation problem, we use a Breusch-Godfrey (BG) test with a 10 lag length. The p-value of the BG test is near zero (2.1e−10), so we reject the null hypothesis of no serial correlation, and we conclude that there is a serial correlation problem in our VAR model ($\chi^2 = 123.1729, df = 40$).

If serial correlation is found in the VAR model, we need to increase the original lag order of the VAR model from 14 to 15 and set up a new model with a new lag order, 15, and again check for serial correlation. We must then repeat these steps, increasing the lag order each time, until we find a VAR model with no serial correlation problem. In our case, adding 9 lags to the original lag order creating a VAR model with 23 lags, thereby removing the serial correlation problem. The resulting BG's p-value is 0.2283. We also confirm the non-existence of serial correlation in VAR(23) by using BG lag orders of 9, 8, 7, 6, 5. Moreover, we make sure that the VAR model with the optimal lag, VAR(23), is stable. The stability test shows that both CADOWD and DEM2008 are stable, with root stabilities of 1.00002 and 1.000109, respectively (both are greater than one). Thus, after investigating the serial correlation issue and stability of the VAR model, we can take the lag order of 23 as the VAR optimal lag order, k.

The significant coefficients in VAR(23) show that DEM2008.l13 (lag 13 of DEM2008) has the most-significant effect on CADOWD. Also, CADOWD.l2 (lag 2 of CADOWD) has the most-significant effect on DEM2008.

The coefficient of lag 13 of DEM2008 is 0.00747, meaning that the prediction market has the most significant affect on the financial market, with 13 lags. In other words, since the time series has 5-minute intervals, those 13 lags are equal to 65 minutes; so we can say, therefore, that the Canadian financial market incorporates the U.S. prediction market data after about one hour. Moreover, since we use a logarithmic transformation of the time series, the coefficient of DEM2008.l13 means that a 1 percent increase in DEM2008 results in a 0.00747 percent increase in CADOWD.

To understand the causality of this relationship, we use Wald testing. Wald test statistics for DEM2008 $\rightarrow$ CADOWD and CADOWD $\rightarrow$ DEM2008 are 33.974 and 33.374, respectively. The p-values related to both test statistics are 0.066 and 0.075, which are both greater than 0.05% and show that we cannot reject the null hypotheses, "DEM2008 does not Granger cause CADOWD" and "CADOWD does not Granger cause DEM2008". Thus, we conclude that there is not a significant causal relationship between DEM2008 and CADOWD for the time series CADOWD-DEM2008-M5-1.

As a final step, to confirm the results of a significant relationship, we need to make sure that there is no structural break in either the CADOWD or DEM2008 time series. For this purpose, we search for any structural break using a Zivot-Andrew (ZA) test. The results of ZA testing show a possible structural break in the DEM2008 series, hence we must avoid claiming that the existent bilateral relationship is significant. Then, we confirm that there is no significant causal relationship when we consider the existence of structural break in CADOWD-DEM2008-M5-1.

As we continue the TY procedure for subsequent iterations, the ZA test might show different results for a possible structural break because in each iteration of the rolling day strategy we are omitting observations from the beginning of the time series and we may pass the break point.

We continue testing the TY procedure up to the last iteration, CADOWD-DEM2008-M5-452. The summery of 452 iterations of TY procedure shows that we do have more significant causal relationships at the end of iterations, when we approach approach the election date, which means that the prediction market starts to impact global financial market near election date more frequently. Also, by comparing the summary of CADOWD-REP2008-M5 iteration results, we can see that the REP2008 has more significant causal relation ship at the end, which implies that causal relationships between CADOWD and

REP2008 are stronger than those between CADOWD and DEM2008 for the same two pairs of 5-minute granularity time series.

We extend our testing process by generating new time series with different granularities described in Section 2.4.3. For example, Figure 2.6 shows the causality directions of PM→SM, SM→PM, SM↔PM, and "No Causality" for the pair CADOWD-REP2008. This figure has 11 bars, which are related to the time series of 5, 10, 15 and 30-minute granularities, and 1, 2, 4, 8, 12 and 16-hour granularities. The last bar is related to the daily time series.

In Figure 2.6, each bar shows the causality directions of a specific granularity. In each bar, blue shows the causality direction from DEM2008 to CADOWD (PM→SM), yellow shows causality direction from CADOWD to DEM2008 (SM→PM), green, which is a mixture of blue and yellow, shows bilateral direction (PM↔SM), and finally, red indicates no causality. The white gaps in each bar show that the TY procedure was not tested on those time series. The bars are not aligned on the right side of Figure 2.6, because iterations are stopped at different time points based on time series granularity, to make sure that we have enough observations to run TY procedures when approaching the election date. The white gaps in the middle of each bar are due to missing data for some specific dates, mostly holidays, which cause some iterations of the TY procedure to skip over those dates.

Figure 2.6: Causality direction of CADOWD-REP2008

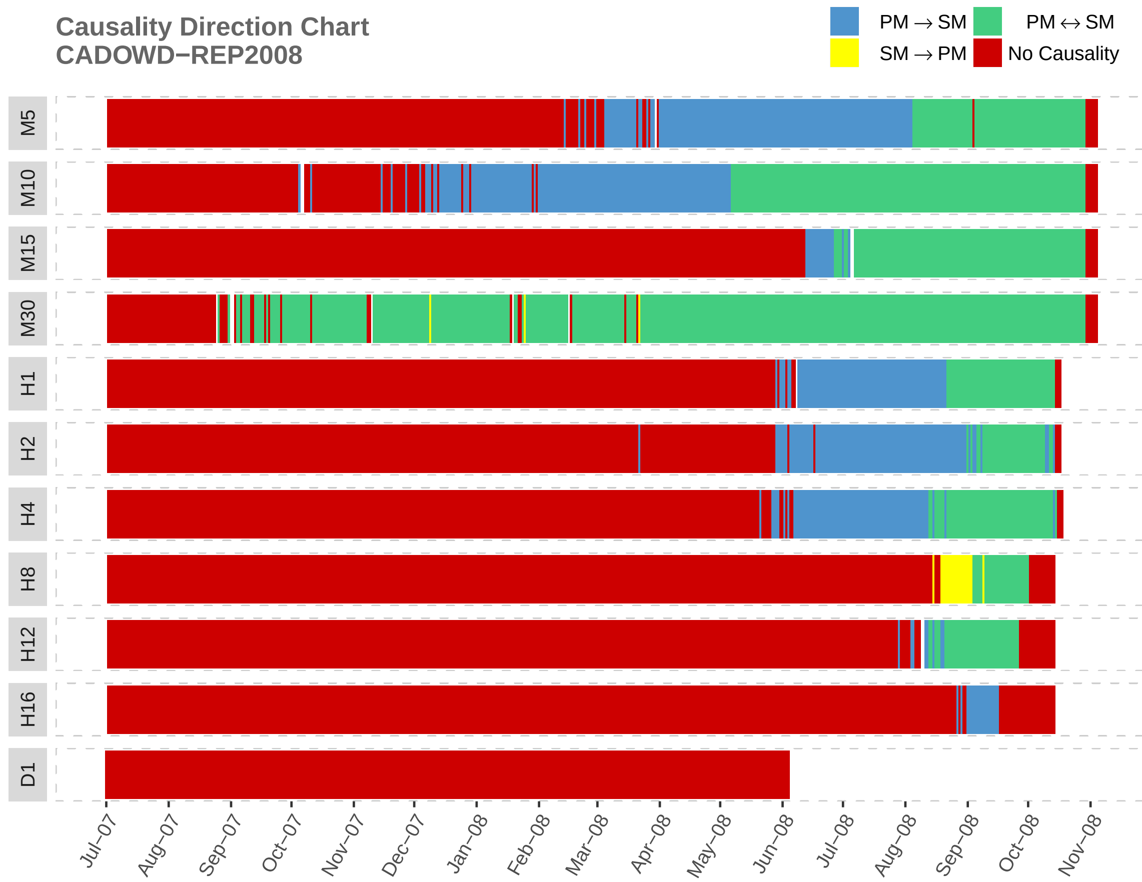


Table 2.2 summarizes the test results presented in Figure 2.6. Numbers in the middle of the table represent the percentage of significant causal relationships in the selected series, direction and granularity. In other words, this table shows the percentage of unique causal relationships for PM→SM and SM→PM. The overlaps of significant causal relationships are shown in the "Both" direction, and the percentage of non-causality is shown in the "Non" row. For each Series, the numbers in each column sum to unity. For example, the first row of Table 2.2 shows that the Canadian financial market incorporated information on the Republican prediction market at 5-minute and 10-minute intervals, with 31% and 32% significant causal relationships, which is better than the results for other granularities.

(Numbers are rounded to two decimal places)

Table 2.2: Percentage of significant causality for each direction (CADOWD-REP2008)

Direction	M5	M10	M15	M30	H1	H2	H4	H8	H12	H16	D1
PM→SM	0.31	0.32	0.03	0.00	0.16	0.21	0.15	0.00	0.02	0.04	0.00
SM→PM	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.04	0.00	0.00	0.00
Both	0.18	0.36	0.25	0.82	0.12	0.08	0.14	0.06	0.09	0.00	0.00
Non	0.52	0.32	0.72	0.17	0.72	0.71	0.71	0.90	0.88	0.96	1.00

Table 2.3 shows the coefficient means for PM→SM causal relationships for different granularities.

Table 2.3: Coefficient's mean for PM→SM, CADOWD-REP2008

	M5	M10	M15	M30	H1	H2	H4	H8	H12	H16	D1
Mean	0.0038	-0.0002	0.0143	0	-0.0109	0.0035	-0.0304	0	0.0238	0	0

Tables 2.4 and 2.5 summarize the information presented in Tables 2.2 and 2.3. Table 2.4 shows the percentage of significant causalities averaged from the rows with 11 granularities represented in Table 2.2. Going forward, we refer to the numbers in Table 3 as the time invariant *Expected Effect* (EE) percentages. Based on the four mutually exclusive causal relationships, described in the research hypothesis (Section 2.3.1), the four Expected Effect percentages are defined as follows: Expected Lead Effect (ELE), Expected Follow Effect (EFE), Expected Feedback Effect (EBE) and Expected Neutral Effect (ENE) percentages.

Table 2.4: Rolling Day Expected Effect percentages

Series Name	PM→SM ELE Percentage	SM→PM EFE Percentage	Both EBE Percentage	Non ENE Percentage
CADOWD-DEM2008	0.005	0.127	0.176	0.692
CADOWD-REP2008	0.113	0.005	0.192	0.690

Table 2.5 shows the sum-products of EE (Table 2.2) and their relevant significant causal relationship coefficient means (Table 2.3) with 11 granularities. The numbers in Table 2.5 represent the *Expected Effect* magnitudes, which have four possible categories depending on whether they indicate Lead, Follow, Feedback or Neutral causal directions.

Table 2.5: Rolling Day Expected Effect magnitudes

Series Name	PM→SM ELE Magnitude	SM→PM EFE Magnitude	Both EBE Magnitude	Non ENE Magnitude
CADOWD-DEM2008	-0.002	-0.158	-0.061	-0.235
CADOWD-REP2008	-0.004	-0.004	-0.006	0.171

We discuss the implications of this section and Section 2.6.2 together in Section 2.6.3.

2.6.2 Moving Window Strategy Results

With the Moving Window strategy, we again perform the TY process 470 times for the CADOWD-DEM2008 pair, using a window width of 5 days. We have selected the window widths based on time series granularity. Time series with finer granularities are tested with smaller window widths and series with coarser granularities are tested with larger window widths. For example, a 5-minute (M5) time series is tested with four different windows, which are 3, 5, 10 and 15 days. Table 2.6 shows the window widths for different granularities.

Table 2.6: Moving Window widths based on time series granularity

Window	M5	M10	M15	M30	H1	H2	H4	H8	H12	H16	D1
3 Days	✓	-	-	-	-	-	-	-	-	-	-
5 Days	✓	✓	✓	-	-	-	-	-	-	-	-
10 Days	✓	✓	✓	✓	✓	✓	-	-	-	-	-
15 Days	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-
30 Days	-	-	-	-	-	-	✓	✓	✓	✓	-
60 Days	-	-	-	-	-	-	-	-	-	✓	✓

We have repeated the Moving Window strategy with the various window widths listed in Table 2.6. Tables 2.7 and 2.8 summarize the percentage of significant causal relationships (*Expected Effect* percentages) and the sum-products of Expected Effect with the relevant coefficients (*Expected Effect* magnitudes) for the pair CADOWD-DEM2008 in the four directions for all window widths.

Table 2.7: Window Strategy Expected Effect percentages, all granularities

CADOWD-DEM2008 Window	PM→SM ELE Percentage	SM→PM EFE Percentage	Both EBE Percentage	Non ENE Percentage
3	0.020	0.018	0.089	0.873
5	0.029	0.028	0.135	0.809
10	0.028	0.041	0.217	0.713
15	0.028	0.040	0.245	0.687
30	0.041	0.040	0.311	0.608
60	0.008	0.026	0.378	0.588

Table 2.8: Window Strategy Expected Effect magnitudes, all granularities

CADOWD-DEM2008 Window	PM→SM ELE Magnitude	SM→PM EFE Magnitude	Both EBE Magnitude	Non ENE Magnitude
3	-0.0005	-0.0048	0.0000	0.0033
5	0.0026	-0.0168	-0.0003	-0.0655
10	-0.0028	-0.0279	-0.0063	-0.0840
15	-0.0048	-0.0594	-0.0106	-0.0629
30	-0.0023	-0.0049	-0.0025	-0.0061
60	-0.0019	0.0015	0.0002	-0.0015

Table 2.9 shows the average of each direction over all windows (columns in Table 2.7). Similar to the results of Rolling Day strategy, the first data column, PM→SM, shows that the probability of the Republicans winning rather than the Democrats has more causality effect on the Canadian financial market under Moving Window strategy.

Table 2.9: Window Strategy Expected Effect percentages, all windows, all granularities

Series	PM→SM ELE Percentage	SM→PM EFE Percentage	Both EBE Percentage	Non ENE Percentage
CADOWD-DEM2008	0.026	0.032	0.229	0.713
CADOWD-REP2008	0.058	0.033	0.192	0.718

Table 2.10: Window Strategy Coefficient Effects, all windows, all granularities

Series Name	PM→SM ELE Magnitude	SM→PM EFE Magnitude	Both EBE Magnitude	Non ENE Magnitude
CADOWD-DEM2008	-0.0016	-0.0187	-0.0032	-0.0361
CADOWD-REP2008	0.0006	0.0001	-0.0029	0.0039

We next move to discuss the implications of our results in this section along with Section 2.6.1

2.6.3 Comparing Two Strategies

Table 2.11 shows the ELE percentages and magnitudes presented in Tables 2.4, 2.5, 2.9 and 2.10, related to the CADOWD series for both strategies.

Table 2.11: Expected Lead Effect (ELE) percentages and magnitudes, both strategies

Series Name CADOWD	DEM2008→SM ELE Percentage	REP2008→SM ELE Percentage	DEM2008→SM ELE Magnitude	REP2008→SM ELE Magnitude
Rolling Day Strategy	0.5	**11.3**	-0.002	**-0.004**
Window Strategy	2.6	**5.8**	**-0.0016**	0.0006

Comparing the ELE for the two parties shows that, under both strategies, the percentages and magnitude of the Republican's ELEs are greater than the Democrat's (REP2008→SM). Thus, we conclude that the Canadian financial market tends to incorporate the probability of the Republicans winning, and the probability of the Republicans winning tends to have a causal effect on the Canadian financial market more frequently and strongly. In the next section we move to discuss our results from applying our analysis to all 27 regions and countries.

2.7 Results and Discussion

In the various tables and graphs, presented in the last section, we show the TY procedure tests results for the Canadian market, pairs CADOWD-DEM2008 and CADOWD-REP2008, at different granularities. We conclude that the probability of the Republicans winning rather than the Democrats had more significant causal effects on the Canadian financial market than the reverse would have.

We extend the TY procedure testing for the other 26 financial markets. Tables 2.12

and 2.13 show the result of Rolling Day and Moving Window strategies tested for pairs of the two prediction markets' data and the 27 global financial markets time series, including Canada. By comparing ELE magnitudes, the last two columns of Tables 2.12 and 2.13, we determined which party had more effect on the selected financial markets (in both tables the greater Expected Lead Effect value is indicated by $\uparrow$ in each row).

To investigate whether the global financial markets are informationally efficient with reference to the U.S. election outcome, we can evaluate the existence, magnitude and direction of causal relationships. If there is no considerable Lead Effect (PM $\rightarrow$ SM), we conclude that the financial market is efficient.

Table 2.12: Rolling Day strategy, Expected Lead Effect (ELE) percentages and magnitudes

Countries/Regions Name	DEM2008→SM ELE Percentage	REP2008→SM ELE Percentage	DEM2008→SM ELE Magnitude	REP2008→SM Magnitude
Americas (A1DOW)	1.3	13↑	-0.001	-0.002↑
Latin America (A3DOW)	2.9	7.6↑	0.009	0.037↑
Latin America ex-Mex (A4DOW)	1.6	6↑	-0.004	0.048↑
Austria (ATDOWD)	1.8	3↑	-0.025↑	0.018
Australia (AUDOWD)	7.8	13.2↑	-0.007	0.034↑
Belgium (BEDOWD)	9.8↑	7.5	-0.001	-0.02↑
Brazil (BRDOWD)	1.6	9.7↑	0.009	-0.016↑
Canada (CADOWD)	0.5	11.3↑	-0.002	-0.004↑
Germany (DEDOWD)	5.9↑	1.2	-0.013↑	-0.006
Denmark (DKDOWD)	7.4↑	2.9	0.003	-0.008↑
Europe (E1DOW)	11↑	5.4	0.002	0.023↑
Europe ex-U.K. (E2DOW)	8.7↑	4.2	0.001	0.018↑
Europe - Nordic (E3DOW)	8.4↑	3.7	-0.036↑	0.014
Finland (FIDOWD)	9.5↑	4.6	-0.045↑	0.019
France (FRDOWD)	10.8↑	3.1	-0.001	-0.011↑
Greece (GRDOWD)	2.4	3.5↑	0	0.003↑
Hong Kong (HKDOWD)	3.9↑	2.7	-0.001	0.005↑
Indonesia (IDDOWD)	0	2.1↑	0	0.015↑
Ireland (IEDOWD)	6.6↑	6.6↑	-0.006	-0.071↑
Italy (ITDOWD)	8.2↑	6.8	0.002	0.055↑
Japan (JPDOWD)	3.2	6.2↑	0	-0.002↑
Mexico (MXDOWD)	1.5	3.1↑	0.001	-0.017↑
Malaysia (MYDOWD)	0.9	23.9↑	0	0.013↑
Netherlands (NLDOWD)	11.2↑	6	-0.012↑	-0.009
Norway (NODOWD)	3.6↑	2.9	-0.166↑	0.005
New Zealand (NZDOWD)	1.4	5.2↑	0.001	0.014↑
China (DJCHINA)	0.5	7↑	0↑	0↑

Table 2.13: Moving Window strategy, Expected Lead Effect (ELE) percentages and magnitudes

Series Name	DEM2008→SM ELE Percentage	REP2008→SM ELE Percentage	DEM2008→SM ELE Magnitude	REP2008→SM Magnitude
Americas (A1DOW)	2.9	4↑	-0.0001↑	-0.0001↑
Latin America (A3DOW)	2.9	5.7↑	0.0001↑	-0.0001↑
Latin America ex-Mex (A4DOW)	3.9↑	3.8	0↑	0↑
Austria (ATDOWD)	2.6	2.8↑	0.0003↑	0
Australia (AUDOWD)	3.2↑	2.8	0.0036↑	0.0002
Belgium (BEDOWD)	2.6	3.2↑	0↑	0↑
Brazil (BRDOWD)	3.7	4.1↑	-0.0013↑	0
Canada (CADOWD)	2.6	5.8↑	-0.0005	0.0033↑
Germany (DEDOWD)	3.5↑	3.5↑	0.0007↑	0
Denmark (DKDOWD)	1.9	3.3↑	-0.0003↑	0
Europe (E1DOW)	3	3.6↑	0↑	0↑
Europe ex-U.K. (E2DOW)	3.1	3.7↑	0.0012↑	0
Europe - Nordic (E3DOW)	3.7	4.1↑	0↑	0↑
Finland (FIDOWD)	3.2↑	2.7	0↑	0↑
France (FRDOWD)	3.5↑	2.7	0↑	0↑
Greece (GRDOWD)	3.3↑	2.8	0	0.0001↑
Hong Kong (HKDOWD)	2.1↑	1.7	-0.0011↑	0
Indonesia (IDDOWD)	1.5	3.7↑	0↑	0↑
Ireland (IEDOWD)	1.7	2.8↑	0↑	0↑
Italy (ITDOWD)	2.7	2.8↑	-0.0045↑	0
Japan (JPDOWD)	1.7↑	1.5	0↑	0↑
Mexico (MXDOWD)	3.3	5.5↑	0.0011↑	-0.0007
Malaysia (MYDOWD)	2.7	3.2↑	-0.0033↑	0
Netherlands (NLDOWD)	3.3	3.7↑	0.0004↑	0.0002
Norway (NODOWD)	2.9	3.2↑	0	-0.0001↑
New Zealand (NZDOWD)	2.2	2.9↑	0.0004↑	-0.0001
China (DJCHINA)	1.6↑	1.1	0↑	0↑

Comparing the behaviour of each of the 27 financial markets under the two testing strategies (using results shown in Tables 2.12 and 2.13), in most cases shows that the two strategies identify the same prediction market as most likely to affect a financial market. For example, the Canadian financial market, CADOWD, was affected by the probability of the Republican party winning under both strategies, and the Expected Lead Effect percentage (ELE percentage) of Rolling Day strategy, 11.3%, is greater than the ELE percentage of the Moving Window strategy, 5.82%. Thus, the Canadian financial market tends to respond to the U.S. election outcome in longer rather than shorter time horizons. The ELE percentage and ELE magnitude presented in Tables 2.12 and 2.13 are not comparable since the Rolling day strategy examines the long-term causality relationships, and the Moving window strategy examines short-term causality relationships.

Tables 2.12 and 2.13 show that the average ELE percentages of the 27 countries or regions are 3.49% and 8.14% for the Moving Window and Rolling Day strategies, respectively. Under Moving Window strategy, most of the ELE percentages are less than 5% (average 3.49%), which implies short-term financial-market efficiency. However, considering our 95% confidence interval, any ELE percent illustrated in Table 2.13 that is at or less than 5% might happen by chance, given the larger number of time-series pairs we consider.

In almost all cases, except Mexico and Indonesia, the ELE percentages of the Rolling Day strategy are much greater than the ELE percentages of the Moving Window strategy. This fact generally implies that the financial markets tend to incorporate the U.S. election outcome in the long term. On the other hand, Mexico and Indonesia's financial markets tend to incorporate the U.S. election outcome in the short term. However, under the Moving Window, the ELE magnitudes of Mexico and Indonesia are smaller than the Rolling Day ELE magnitudes.

Considering long-term effects under the Rolling Day strategy, the ELE percentages of all three European indices –E1DOW, E2DOW and E3DOW– confirm that the European markets are frequently influenced by Democratic party wins, however, the ELE magnitudes of all three European indices show that the Republican party had a more powerful effect. Thus, we conclude that European markets are not highly efficient in respect to the U.S. election outcome (ELE percentages average 9.36%).

Also, regarding all three Americas' (North, South and Central) indices –A1DOW, A3DOW and A4DOW– the results show that North and South American countries' financial markets are more likely to be influenced by the Republican party than are central ones. Moreover, Americas' ELE percentages average 8.8%, showing that financial markets are not highly efficient in any of the Americas regarding U.S. presidential election information.

2.8 Conclusions

The primary objective of this research was to investigate the existence and direction of causality relationships between the U.S. presidential election outcome and various global financial markets. For this purpose, we have implemented the Toda–Yamamoto causality test to overcome the pitfalls of conventional Granger causality testing. We estimated VAR models with additional lags, and choose a modified version of Wald testing to ensure asymptotic χ^2 distribution.

This research has taken two different strategies for estimating VAR models: Rolling Day and Moving Window. The first considers longer time horizons, tested in multiple iterations, and the second examines shorter time horizons. Using two strategies allowed us to investigate the Efficient Market Hypothesis (EMH) through two different lenses.

In almost all cases, the ELE magnitudes of both Rolling Day and Moving Window strategies show that the Republicans had more powerful effects on financial markets.

The semi-strong form of EMH, in this research, implies that changes in financial-market prices are independent from changes in the probability of a certain party winning the U.S. election. In other words, financial market prices incorporate all publicly available information, including that related to a U.S. election outcome.

Financial markets can be divided into three main categories: highly efficient, efficient, and moderately efficient. The Rolling Day results show that, in the longer term, some countries such as Austria, Greece, Hong Kong, Indonesia, Mexico and Norway are highly efficient since their Expected Lead Effect percentages are less than 5%. Countries or regions such as Latin America, Latin America ex-Mex, Belgium, Brazil, Germany, China, Denmark, Europe ex-UK, Europe-Nordic, Finland, Ireland, Italy, Japan and New Zealand are efficient since their Expected Lead Effect percentages are between 5% and 10%. Finally, with Expected Lead Effect percentages greater than 10%, countries or regions such as the Americas, Australia, Canada, Europe, France, Malaysia and the Netherlands are moderately efficient.

Finally, several questions remain to be answered. For instance, why the relationships are what we observe is something for future research as there may be many drivers behind the observed results. Such drivers are geopolitical, social, financial, political, etc. that are far too broad for us to capture with our study.

Chapter 3

Determining the impact of different IP protection policies of YouTube videos on YouTube channels

3.1 Introduction

Digital content producers report that piracy costs could reach U.S.$4.2 trillion by 2022, jeopardizing 5.4 million jobs (ICCBelgium, 2017). This is a major issue as 45% of all music is digital, and is growing faster than other formats; in 2015 total music-industry revenues grew by 3.2 per cent to U.S.$ 15.0 billion, and digital revenues rose 10.2 percent to U.S.$ 6.7 billion (IFPI, 2016). Similar to the music industry, all digital content-producers' gains are in danger due to the extensive consumption of copyright-infringing materials (i.e., any copyright-protected work that is reproduced, distributed, or made into a derivative work without the permission of the copyright holder). Misapplication of some legislation, such as the "safe harbor" provision defined in section 512 of The Digital Millennium Copyright Act (DMCA), allows some service providers, like YouTube, to bypass the standard rules of licensing. For example, the 2015 Digital Music Report argues that YouTube is misusing the DMCA safe harbor act to avoid paying its fair share of music licensing fees, and YouTube should not qualify for safe harbor IFPI (2015).

This misapplication is surprising as the act implements two 1996 World Intellectual Property Organization (WIPO) treaties (Office, 1998).

Nowadays, over two billion active YouTube users watch one billion hours of video per day (YouTube, 2019; videonitch, 2017). YouTube users upload over 500 hours of user-generated content (UGC) every minute and watch 46000 years of video content each year (Chi, 2019; Iqbal, 2019). Generally, UGC refers to any form of digital content, such as videos, music, text or images, that is created and publicly shared by end-users of an online system or a website (e.g., any video created and uploaded by YouTube's users). Of the 500 hours of video uploaded every minute, there is an abundance of copyright infringing materials uploaded resulting in YouTube paying US$ 1.25 billion to copyright holders since it established a reimbursement scheme in 2007 (statisticBrain, 2017).

While YouTube earned US$ 15.15 billion in 2019, Google websites reportedly earned US$ 113.26 billion (Spangler, 2020; Clement, 2020). At the same time, YouTube pays creators, to create content to post on YouTube in order to attract and retain viewers. For example, PewDiePie, a Swedish comedian generating content for YouTube with over 103 million subscribers, was paid US$13 million in 2019 (Berg, 2019).

Based on a 2016 YouTube official report, since 2007, two billion dollars has been paid to content-owners who claim and monetize their video material used by others. As of July 2015, more than 400 million videos are claimed by 8,000 plus YouTube partners using the proprietary YouTube content ID system. The content ID system automatically searches, identifies, and claims any newly uploaded YouTube videos against a catalog of content posted by copyright owners. There are more than 50 million active reference files (content posted by copyright owners) in the content ID system database (YouTube, 2019).

Despite its well intentioned origins, the content ID system has a non-zero false-positive identification probability. For example, some critics point out that the content ID system penalizes creators who legally use copyrighted assets in their videos. Also, the automated matching system can suppress fair uses of copyrighted content (Tassi, 2013). Moreover, in some cases, the content ID system may be used as a censorship tool. For example, a video game company might flag an unfavorable review as infringing on content.

Content owners, using the content ID system, may choose one of three different policies against a claimed video (a video that is identified as infringing on the originally created content): Block, Monetize or Track. The Monetize policy, which is the most common policy, selected by 90 percent of content owners (Google, 2008), keeps the infringing video on the YouTube website, but the advertising earnings generated by the claimed video go to the copyright-holder's account. The Block policy prevents the claimed video from being uploaded. The Track policy allows the claimed video to be uploaded and the copyright-holder may see all statistics collected by YouTube for the claimed video.

The main objectives of this research is to help YouTube partners and content owners to answer two questions : 1) what is the impact on a copyright-holder's channel (the collection of all videos posted by the copyright-holder) viewership from a claimed video? 2) does the impact of a claimed video on a copyright-holder's channel viewership change if the copyright-holder uses a different IP protection policy?

In this research, we use panel data analysis to compare the effect of two different IP protection policies, Monetize and Track, on YouTube channel viewership. User-generated content (claimed videos, referred to as UGC moving forward) viewership data is collected from two different YouTube channels, Monetizer and Tracker, which use Monetize and Track IP protection policies, respectively.

In the remainder of this introduction, Subsections 3.1.1 and 3.1.2 discuss YouTube IP protection-policy challenges and explain YouTube Content ID System in greater detail. Subsection 3.1.3 provides additional background on digital piracy and how content producers changed business practices to respond to piracy.

3.1.1 YouTube IP Protection Challenges

YouTube, the world's second largest search engine and the world's most popular entertainment and learning tool as measured by site traffic, hosts millions of videos on its servers Davies (2018). All too frequently, video creators use YouTube services to share materials they did not create or have the right to redistribute.

Infringements of copyright laws, i.e., the act of creating user-generated content using others' materials without their consent, causes problems for YouTube. For example, in early 2007, Viacom Inc., a multinational media conglomerate, sued YouTube for US$1 billion, claiming YouTube did not take down over 100,000 copyright infringing videos (Eff, 2017). Viacom argued that YouTube was not entitled to "safe harbor" protection under the DMCA when it shared more than 160,000 audiovisual "clips" that generated 1.5 billion views. Viacom claimed that YouTube was aware of the fact that these clips did indeed infringe on Viacom's copyrights. However, YouTube denied those claims, arguing that the lawsuit threatened the Internet freedom (BBC, 2007*a*).

Based on the Viacom lawsuit arguments, after a few months, several class actions were filed against YouTube on behalf of a group of sports-league and music publishers. The English Premier League claimed that YouTube had "knowingly misappropriated" their copyrighted assets by encouraging people to view UGC on YouTube (BBC, 2007*b*). YouTube has denied being involved in mass copyright infringement, claiming that its activity complies with section 512 of the DMCA, "safe harbor" provision.

YouTube Safe Harbor

The "safe harbor" provision states that internet service providers (ISPs) are not responsible for what their users put on websites. In other words, the DMCA set out the safe harbor provision to protect ISPs from unreasonable liability. However, ISPs are obliged to block or remove infringing content that they are aware of, or for which it receives a valid notice.

DMCA offers four safe harbors that substantially limit the responsibility for copyright infringement: 1) transitory digital network communications, which transmit materials through the system or network without modification of the content, 2) system caching, which temporarily store web content to speedup the delivery of digital content, like Google Cache, 3) information residing on a system or network at the direction of users, like hosting user-generated content on YouTube, and 4) information location tools, like search engines that provide URLs (Cornell, 2017).

Based on the "safe harbor" idea, ISPs are not responsible for actively searching for copyright infringements, but must act if they find any infringing content in their networks. However, ISPs may learn about copyright infringing content either through a take-down notices or by other means. Thus, learning about the existence of copyright infringing content, referred to as "red flag knowledge," legally compels ISPs to take action.

In the case of the YouTube vs. Viacom lawsuit, YouTube won the argument because Viacom had no proof showing "red flag knowledge" of specific infringements. Viacom only argued general knowledge of infringement, which had no legal footing in court (Masnick, 2013). Finally, in March 2014, Viacom and Google settled the lawsuit by releasing a statement. The Viacom lawsuit motivated YouTube to significantly improve its services by creating a state-of-the-art system, called the "content ID system," in late 2007 (Spangler, 2014). This system automatically identifies copyright infringements by scanning all content, to protect YouTube advertising business, copyright-holders' rights and a billion YouTube users, creators and viewers, who are generating billions of dollars in revenue (Google, 2007).

3.1.2 YouTube Content ID System

The YouTube intellectual property management system consists of three major components: 1) the YouTube rights management system, 2) the content ID system, and 3) YouTube videos. The first component identifies the owners of intellectual property (assets) and defines appropriate policies to protect the asset owners' rights. The second component, the content ID system, scans user-generated video content to find any match to copyrighted assets and applies the specified rights policy to the matching videos. In the content ID, system the representation of intellectual property is named a "reference." The last component, YouTube video, can be a public representation of the intellectual property, like a video or audio, that is defined in the system (YouTube, 2017*b*).

YouTube's content ID system utilizes a large capacity scanning system that matches all user-generated content against all assets' fingerprints that exist in the content ID database. The Content ID system database contains digital fingerprints of assets that are claimed by the legal rights holders.

Record companies, game developers, and YouTube partners have access to YouTube's Content Management System (CMS). The CMS helps content-owners to protect their assets by creating references in the Content ID database. Content owners, using the CMS, can associate references with assets, and sometimes an asset can be connected to multiple references. For instance, a re-mix music asset could have multiple references.

The content ID system continuously compares new user-generated content to the references for content-owners' assets. It automatically "claims" any matching videos, and applies the IP protection policy set by the asset owner to the claimed videos. Moreover, it performs a full "legacy scan," which can take up to six months, to identify those matching videos that were uploaded before references were filed in the content ID system database (YouTube, 2017*c*).

3.1.3 Digital Piracy and Music Industry Response

In this section we provide historical context on why there are three different IP protection policies on YouTube. In particular, these IP protection policies are motivated by a shift in the business models used by music distributors.

The Economics Frontier estimates that the size of the global piracy market is in the range of US$250 − $650 billion (Economics, 2011). Illegal downloads negatively affect everything, including profits and job security. For example, analysis by the Institute for Policy Innovations (IPI) shows that the piracy of sound recordings causes U.S.$12.5 billion in economic losses annually and more than 71,000 U.S. job losses (Siwek, 2007). However, there is no consensus on the impact of online piracy on music sales. McKenzie (2009); Oberholzer-Gee and Strumpf (2007) argue that online piracy either has no effect or a positive effect on music sales. In the remainder of this section, we first discuss how the music industry responded to online piracy: by introducing new business models. Then we discuss two of these business models in greater detail. We conclude the section by comparing the new business models using frameworks found in the literature.

Due to online piracy, the music industry has had to respond. One response, which we focus on in this paper, was to change its business model from selling music on compact discs (CDs) to supporting online music stores. The business model transformation also lead to different pricing strategies for music: download-to-own model, subscription-based, and ad-supported (Wikström, 2013). One of the first pricing strategies used is the download-to-own model, in which a consumer pays a one-time fee to download electronic content to save it on a personal device. Apart from the download-to-own model, two other different models exist: subscription-based and ad-supported. For example, some digital music stores, such as iTunes, provide on-demand subscription-based services with which consumers can access all channel content so long as they are members. Other stores, such as YouTube, are streaming ad-supported content, which is free to use and generates advertisement revenue for content owners.

The 2018 year-end music industry revenue report, published by the Recording Industry Association of America (RIAA), shows that streaming services are now the majority source of recorded music revenue. For example, the streaming music platform sales in the U.S. accounted for US$7.4B in revenue, making up 75% the total revenue across all formats for the year 2018. Considering the fact that physical CD revenue was 12% in 2018, we can say that most firms use either subscription-based or ad-supported business models. We will focus on these two model for the remainder of the section.

The two business models, subscription-based and ad-supported, are studied in the literature especially with respect to the music industry. Some authors take a firm perspective (Lin et al., 2012) and others take a consumer perspective (Papies et al., 2011). Regardless of the perspective, all come to the same conclusion: an ad-supported business model is more beneficial than a subscription-based model. Given this conclusion, we empirically study two different ad-supported business models in this paper for the YouTube distribution channel. In particular, it is up to each firm on YouTube to select an intellectual property (IP) protection policy (IPPP). The two IPPPs we consider are: 1) monetize and 2) track. The monetize policy monetizes all content that uses any IP owned by the generating firm (in our case we refer to these firms as a YouTube channel or channel for short). Monetize policy enables copyright holders to have YouTube run ads on copyright infringing UGC videos, and all revenue for these ads is collected by the copyright owner. For simplicity, we omit the case of a UGC video infringing on multiple copyright owners.

The track policy will keep track of all content that uses any IP owned by the generating channel. Under the Track policy, copyright holders are able to monetize only their own original videos. In addition, copyright holders can also collect copyright infringing UGC video data. UGC viewership data allows the copyright holder determine whether UGCs drive people to the original channel's videos or other distribution channels (i.e., iTunes and

Spotify). To our knowledge, no other work compares these two IPPPs on YouTube or any other setting.

One may argue for or against each IPPP. When comparing two different YouTube channels each with a different IPPP, a user generated content (UGC) channel may select the channel with the most lenient IPPP. Doing so will enable the UGC creator to collect all revenue from its generated content, even that using IP from the tracking channel. However, the tracking channel may not have the content the UGC creator wants and so the creator may in turn move to the monetize channel. In either case, a listener or viewer of UGC content may choose to follow up to see the original content. We may expect no difference in consumers of UGC content with respect to following up with the original content. However, if there are differences in how UGCs use the content of each channel, then there may be a difference of what consumers of UGC content do with respect to following up to see/listen to the original content.

In the remainder of the research, Section 3.2 presents a formal problem statement and reviews the relevant literature. Section 3.3 describes our data and discusses data collection methods. Section 3.4 introduces the statistical methods used in our analysis, and Section 3.5 describes the results of our analysis using two panel-datasets collected from two music channels with different IP protection policies –Monetize and Track. Finally, the last two sections conclude the paper.

3.2 Literature Review and Problem Statement

This section first briefly reviews relevant literature, and then it articulates the research problems addressed.

3.2.1 Literature Review

Although choosing appropriate IP protection policies is an integral part of managing digital assets on YouTube, it remains an under-investigated topic in the academic literature. This section considers two streams of literature: 1) digital rights management (DRM) systems and 2) IP protection policies on YouTube. In the remainder of this section we discuss these two streams of literature in greater detail.

Digital rights management (DRM) systems, focuses on rights management tools that use digital technologies. The YouTube CMS is a type of DRM system that empowers channel owner's ability to protect their IP. DRM systems help copyright holders, usually media companies, simultaneously fight digital piracy and monetize their content. DRM systems have emerged from the development of access control technologies for protecting the use of copyrighted materials (Rosenblatt et al., 2002). DRM systems provide copyright protection for most digital content (e.g., music, video, image) across multiple platforms (e.g., mobile phone, personal computer) to preserve owners' rights from illegal distribution (e.g. file sharing websites, compact disc and external memory). As a matter of fact, a DRM system includes various information technology (IT) components that aim to control digital content distribution (Lyon, 2001).

Existing literature on DRM systems falls into two main substreams: 1) definitions and theoretical framework and 2) DRM usage and empirical studies. The first category defines a DRM system's component and describes how all of the components connect and work together (Cope and Freeman, 2001; Buhse, 2001; Fetscherin, 2002; Foroughi et al., 2002; Pitkänen and Välimäki, 2000; Armstrong, 2003). Despite all of the work in this substream, there is no standard definition of a DRM system. However, The Association of American Publishers (AAP) defines DRM systems as the combination of various technologies (software and hardware) and processes that facilitate intellectual property (IP) protection in the digital content. Additionally, major DRM systems providers (e.g., YouTube, IBM, Sony, Microsoft, and Adobe) are continuously incorporating new technologies in their DRM systems. For example, IBM uses the Blockchain technology to preserve the IP rights and royalties in the music industry (IBM, 2015).

The second substream of the DRM systems literature addresses practical issues, such as security and encryption technologies to address unauthorized content copies. For example, Halderman (2002) discuss copy-prevention technologies for music compact discs, and Craver et al. (2001) discuss the challenges of watermarking technologies (a copy-prevention technique) in audio clips. Fetscherin and Schmid (2003) compares the usage of DRM systems in the music, film, and print industries. Chorianopoulos et al. (2005) discusses the music-industry's DRM system limitations and proposes a business model that helps accessing music content in various devices (e.g., cell phone, music player and personal computer). Also, Kwok (2002) suggests a license management model for the digital music industry that enables customers to purchase and access music when they are offline.

Of course DRM systems are designed to protect copyright holders and help the copyright holding firms grow, but it is still controversial issue in digital markets. DRM systems benefit copyright holders, however they bring some controversy to firms using such systems. For example, Foroughi et al. (2002) analyze the challenges of balancing compensating copyright-holders and the rights of end-users to access information. Also, Vernik et al. (2011) investigate the impact of DRM systems on the level of piracy in the music industry. The authors find that offering a DRM-free licence decreases music piracy, however, piracy decrease does not assure a notable rise in a company's profit. Our research contributes to the second substream of the DRM system literature stream. In particular, we investigate the impact of different IP protection policies, offered by the YouTube DRM system, on a YouTube channel's viewership and revenue. We do not contribute to the first substream as we do not design or discuss the components of the YouTube DRM system.

The second stream of literature, copyright on YouTube, is rather limited. Based on a Google Scholar search, the number of academic publications that have “YouTube” and “Copyright” as keywords in their titles was less than 125 by the end of October 2019. Some of these early publications discussed the Viacom-YouTube lawsuit (VerSteeg, 2007; Allen, 2007; Collings, 2008; O’Brien, 2008; Hormann, 2009; Hassanabadi, 2011), and the safe harbor provision of U.S. copyright laws (Patten, 2007; Kim, 2007; Driscoll, 2007; Cloak, 2007). Most recent publications discuss fair use of copyrighted material on YouTube (Hunt, 2007; Kumar, 2008; Collins, 2014; Jung, 2014; Bartholomew, 2014; Boroughf, 2015; Solomon, 2015). However, none of these papers discuss what YouTube IP protection policy is most appropriate for a content creator, a gap we address in our research.

To the best of our knowledge, despite the prominent role of IP protection policies for the digital asset industry, research has not quantified the effect of choosing various YouTube IP protection policies on copyrighted assets. This research is a needed first step to help determine the impact of a UGC on a YouTube channel.

3.2.2 Research Questions

The main research question is: What is the difference in YouTube channel viewership between two different YouTube IP protection policies towards user-generated content (UGC) videos using channel assets? To answer this main question, we consider the following subquestions: 1) How does UGC’s viewership affect Monetizer channel (i.e., a channel that uses the Monetize policy) viewership under the monetizing IP protection policy? 2) How does UGC’s viewership affect Tracker channel (i.e., a channel that uses the Track policy) viewership under the tracking IP protection policy?

The hypotheses on the relationship between UGC and content creator viewership are as follows:

H_1^{TC} Tracker model: *there is a significant relationship between UGC viewership and Chan-*

nel Viewership under the Track policy.

H_1^{MC} Monetizer model: *there is a significant relationship between UGC viewership and Channel Viewership under the Monetize policy.*

We use the results of H_1^{TC} and H_1^{MC} to compare the affect of the two IP protection policies on a content creator's channel. The results of this study will provide YouTube content owners with insights on how to handle their copyrighted content so as to increase revenues and channel viewership.

3.3 Data Description

In this research, we use two different music channels' datasets to compare the channels' IP protection policies. Both channels provide electronic music videos to their subscribers, and their videos are also similar in genres (e.g., pop, hiphop, and jazzhop). The first dataset is collected from the Monetizer music channel, which uses the Monetize policy, and the second one from the Tracker music channel, which uses the Track policy. Neither dataset is publicly available, and the data is procured via the channel-owners.

The Monetizer channel (MC) belongs to an electronic dance music record label. MC signs artists to its label and releases their songs on a variety of channels, YouTube being one. The Tracker channel (TC) is an online platform dedicated to finding and sharing the best "chilled" hiphop, jazzhop and triphop music (a form of electronic music). TC provides high-quality music for free. As of June 2019, MC had uploaded 1,549 music videos, which generated 2.3 billion views from its 7.3 million subscribers. By the same month, TC had uploaded 438 videos, which generated 219.1 million views from its 2.3 million subscribers.

MC's data is collected using the YouTube API. We worked with an industry partner that has access to a Montizer channel's dataset from their Content ID system. We has first collect all active-claim lists containing information of over 3 million UGCs, the set of UGCs on YouTube using MC assets. We then collected daily viewership data for claimed videos (UGCs).

Figure 3.1 consists of three charts. The top chart shows the Monetizer channel's normalized daily viewership, from early February 2017 to mid-June, 2017 [1]. The middle chart shows the average of UGC daily viewership, during the same date range. The bottom chart compares Monetizer channel viewership and UGC viewership. For both top charts, the red line is a smooth line, and the blue line shows a fitted linear regression line.

The Tracker channel data collection process differed from the Monetizer channel. TC provides its music free of charge, but asks YouTubers to cite the TC website on their UGC's description by referencing a link that redirects viewers to the channel. Creators who want to use TC's music need to fill in a licensing form where they agree to credit TC assets in their video descriptions. The list of licensed channels helps the TC to understand where their music is being used. TC has no opportunity to benefit from the content ID system, so we confronted some limitations in trying to collect a comprehensive list of UGCs that use TC assets.

To collect TC's dataset, we received access to the licensed-channel lists, which contains over 4000 YouTube content creators. We scan all videos on the licensed channels, every night at 1:00 a.m., via a public YouTube API, and record the viewership information for videos that contain the TC citation keywords in our dataset. We repeat this process every day.

To make the MC dataset as similar to the TC dataset as possible, we searched through the descriptions of over 3 million UGC videos, and selected those UGCs that contained a URL link or keywords related to the MC. We then created the MC panel dataset based on

[1] We collected these datasets for nearly four months on a daily basis. We were not able to collect more data due to time limits and overhead costs that would have required continual buy-in from our industry sponsor. However, the number of observations collected is large enough for running statistical tests.

that condition.

After collecting the TC and MC datasets, we consider three pre-processing steps to clean the data. First, we trimmed outliers to the 99^{th} percentile and removed one percent of observations from the right tail viewership distribution. Second, we use the logged first-difference transformation of panel datasets to avoid potential cyclical behavior in our datasets. Table 3.1 shows both original datasets summary statistics.

Table 3.1: Datesets summary statistics

Statistic	$N \times T$	Mean	St. Dev.	Min	Max
	Tracker Channel (TC) dataset				
Channel Viewership ($CV_{i,t}$)	155,304	323,089.1	146,874.2	94,221	874,510
UGC Viewership ($UV_{i,t}$)	155,304	249.187	4,014.720	1	483,052.000
	Monetizer Channel (MC) dataset				
Channel Viewership ($CV_{i,t}$)	1,107,542	1,499,370	105,096.1	1,261,469	1,810,254
UGC Viewership ($UV_{i,t}$)	1,107,542	124.241	2,940.029	1	392,707

Channel Viewership ($CV_{i,t}$) is the dependent variable; and UGC Viewership ($UV_{i,t}$) is the independent variable, where t ($t = 1, ..., T$) is the time index (Day); and i ($i = 1, .., N$) is the individual index (UGC ID). Both the MC and the TC panel datasets have have relatively large N and T ($N \times T$ is the number of total observations for each datasets). The TC dataset contains 99 days of observations (T) for 5704 UGCs (N). Also, N and T for the MC dataset are 61497 and 110, respectively.

Figure 3.1: Monetizer channel dataset graphs

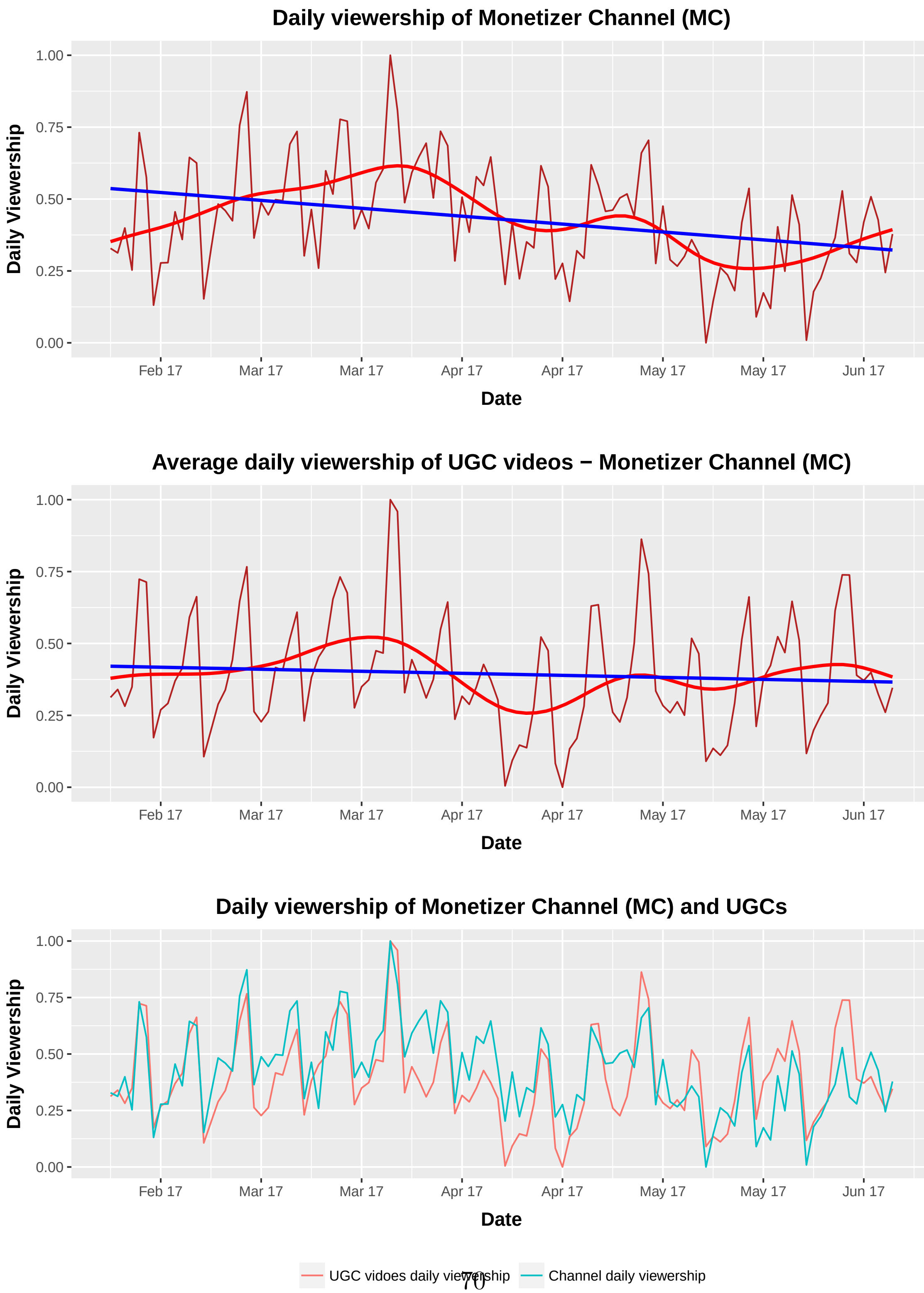
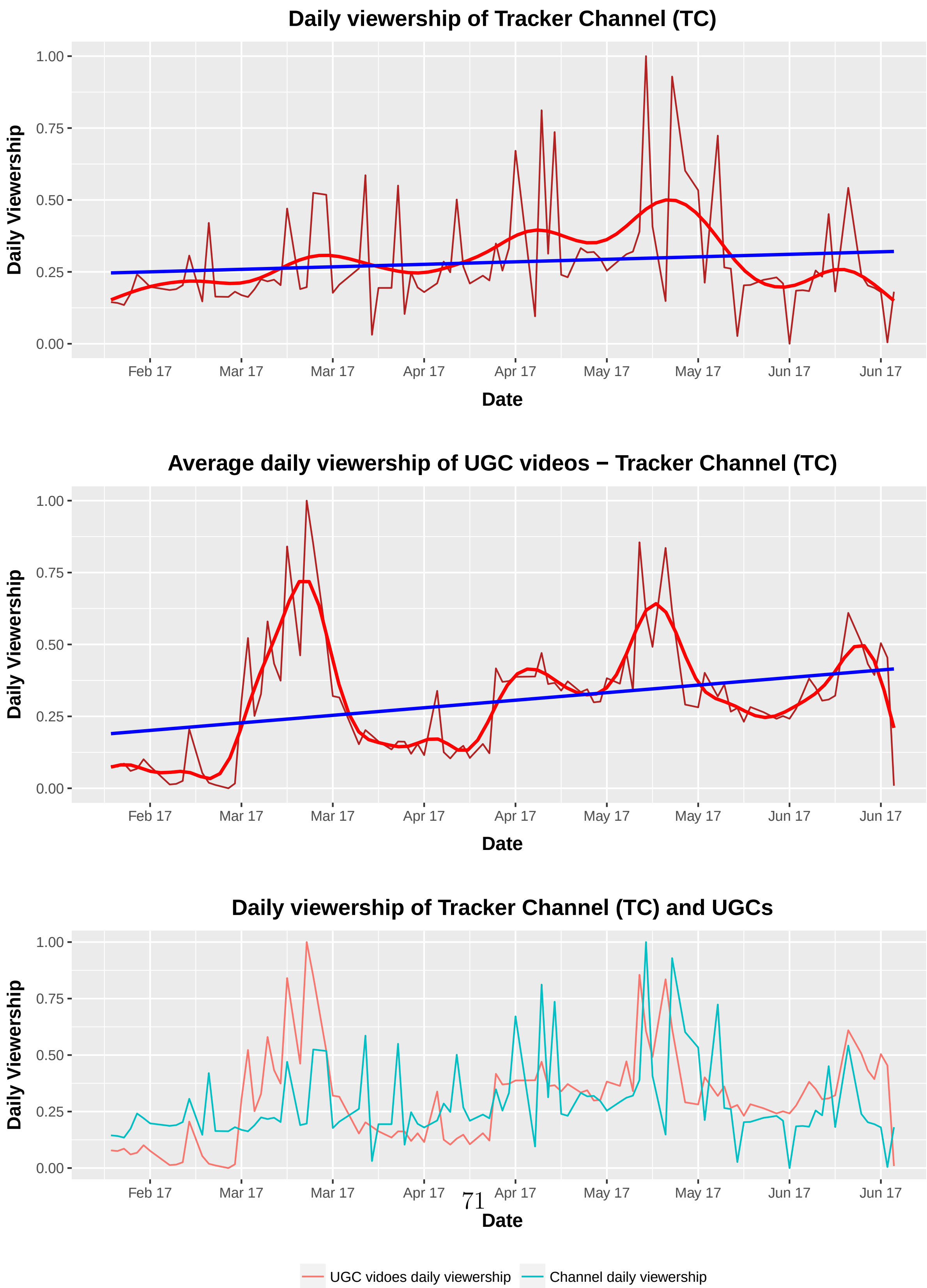


Figure 3.2: Tracker channel dataset graphs



3.4 Statistical Methods

In this research, using the panel datasets described in Section 3.3, we apply panel data analysis to determine the functional relationship between the number of views of the MC (TC) as a function of the number of user-generated videos that use MC (TC) assets.

In this section, first, we briefly discuss panel data analysis advantages, and then introduce the panel data models, estimation, and selection. Finally, we discuss our testing procedure and validation process.

3.4.1 Panel Data Analysis

Panel data analysis is designed for longitudinal time-series data for a fixed set of data generators (individuals or UGCs in our study). Panel datasets typically have two dimensions: a time series dimension, t, and a cross-sectional dimension, i. A panel dataset with $N \times T$ observations involves N units of cross-section data, where $i = 1, \ldots, N$, over T time periods, where $t = 1, \ldots, T$. In other words, in panel datasets, individuals are repeatedly sampled at different points in time. The term "Individual" may represent a city, a firm, a physical person or any kind of entity, in our case, UGC YouTube videos.

There are two types of panel datasets: balanced and unbalanced. In a balanced panel dataset we have the same time periods, $t = 1, \ldots, T$, for each cross-section observation; then the number of observations for a balanced panel is $N \times T$. In contrast, in an unbalanced panel dataset, the number of time periods, T, is not the same for all individuals, i.

Panel datasets may be divided into three categories: short, long, and mixed. The short panel has many individuals and few time periods (large N and small T). The long panel contains many time periods and few individuals (large T and small N). The mixed panel, or macro panel, which is a mixture of short and long, has many time periods and many individuals (Cameron and Trivedi, 2010). Considering an appropriate number of individuals in a panel dataset is important, because a small N or a large N may cause Type I and II errors, respectively. Moreover, based on Greene's (2008) arguments, panel data can be fixed or rotating. A fixed panel observes the same individuals for each period. While a rotating panel contains a set of individuals that change from one period to the next. In this research, we use unbalanced macro panel datasets.

Panel data Analysis Advantages

Panel data allows us to conduct inter-group and intra-group analysis. Wooldridge (2010) argues that solving the omitted variables problem is the fundamental motivation of using panel data analysis. Omitted variables cause heterogeneity bias, which is related to an unobserved property of a variable, in a regression model. Thus, panel data analysis helps researchers construct and investigate variable relationships while eliminating the impact of missing variables and controlling for heterogeneity. Panel datasets combine cross-sectional data, which reflect long-run behavior, and time series data, which contain short-run effects and allows one to measure effects that are not detectable in other datasets. In other words, panel data analysis helps control the consequences of missing or unobserved variables (Hsiao, 2007, 2003).

Hsiao (2007) classified panel data analysis benefits into three main groups: inference precision, computation simplicity and capturing the complexity of human behaviour. Hsiao et al. (1995) argue that panel data analysis is more accurate for model-parameter inference than time series or cross-sectional data analysis, because panel datasets have more degrees of freedom by incorporating cross-sectional and time series data simultaneously. Thus, we have less multicollinearity, which is related to a possible correlation of independent variables in a regression model, and more variation in the data, which improves the efficiency of estimates. In other words, a high correlation between explanatory variables, or a multicollinearity problem, is less likely when variables are changing in two dimensions across time and individuals (Hsiao, 1985).

Panel analysis also helps researchers handle the issue of nonstationary time series, in which the distributions of estimators such as least-squares or maximum likelihood are no longer normally distributed (Phillips and Durlauf, 1986). In panel data in which observations among cross-sectional units are independent, the limiting distributions of many estimators remain asymptotically normal (Binder et al., 2005).

Panel Data Models

Panel data models describe an individual's attributes both across time and across individuals by examining individual effects, time effects, or both, to deal with heterogeneity or unobserved variables. There are two types of panel data models: the pooled model and the individual-specific effects model. The latter can be a fixed effects or random effects model.

The general form of a panel data regression model with two dimensions in which the dependent variable, y, is explained by, K exogenous variables, $x_1, \ldots, x_K$, and non-observable random term, u, for individual i at time t is as follows:

$$
\begin{aligned}
y_{it} =& \beta_{1it} x_{1it} + \beta_{2it} x_{2it} + \ldots + \beta_{Kit} x_{Kit} + \varepsilon_{it} \\
y_{it} =& X'_{it} \boldsymbol{\beta}_{it} + \varepsilon_{it}
\end{aligned}
\tag{3.1}
$$

where X'_{it} is the $(1 \times K)$ row vector of independent variables, and $\boldsymbol{\beta}_{it}$ is the $(K \times 1)$ column vector of coefficients. To consider a constant term in equation (3.1), assume that $x_{1it} = 1$ for all i and t, and the equation can be rewritten as $y_{it} = \beta_{1it} + \widetilde{X}'_{it} \widetilde{\boldsymbol{\beta}}_{it} + \varepsilon_{it}$, where $\widetilde{X}'_{it}$ contains $K-1$ independent variables and $\widetilde{\boldsymbol{\beta}}_{it}$ contains $K-1$ coefficients.

The pooled model, which is the most-restrictive panel data model, specifies constant coefficients, $\boldsymbol{\beta}$, in the general form of a panel data regression model as follows:

$$y_{it} = \alpha + X_{it}'\boldsymbol{\beta} + \varepsilon_{it} \tag{3.2}$$

A constant coefficient assumption, which is very common in cross-sectional analysis, is not commonly used in the panel data analysis literature. The pooled model assumes that the individual effect, α_i, is equal to zero, but if in the real world $\alpha_i \neq 0$, then some of the OLS core assumptions may be violated. For example, homoscedasticity, which means the error terms have the same variance, will not be valid. Panel data analysis using the individual-specific effects models addresses this issue.

In the individual-specific effects model, we assume that α_i captures unobserved heterogeneity across individuals (when $\alpha_i = \alpha$, then equation (3.3) is similar to (3.2)). The general form of an individual-specific effects model is as follows:

$$y_{it} = \alpha_i + X_{it}'\boldsymbol{\beta} + \varepsilon_{it} \tag{3.3}$$

where α_i represents individual-specific effects. If intercepts, α_i, are correlated with independent variables, we are dealing with a fixed effects model. Otherwise, if they are not correlated, the panel model is a random effects model.

In the fixed effects model, we assume that individual-specific effects, α_i, are correlated with the independent variables, X_{it}'. Each individual has a different intercept term, α_i, and the same slope parameters, $\boldsymbol{\beta}$. In other words, α_i represents the leftover variation that cannot be explained by the independent variables.

In the random effects model, we assume that individual-specific effects, α_i, are distributed independently of the independent variables, X_{it}' . In this case, the individual-specific effects can be included in the error term, and equation (3.3) is rewritten as follows:

$$y_{it} = X'_{it}\boldsymbol{\beta} + \varepsilon_{it} \tag{3.4}$$

where $\varepsilon_{it} = \alpha_i + \epsilon_{it}$ (ϵ_{it} is remainder components or a "traditional" error term), and $var(\varepsilon_{it}) = \sigma^2_\alpha + \sigma^2_\epsilon$. Also, $cov(\varepsilon_{it}, \varepsilon_{is}) = \sigma^2_\alpha$.

Panel Data Estimation

In the classical linear regression model, we assume that the regression model is acceptable if independent variables are exogenous, that is not affected by the other variables in the model, and errors are uncorrelated and homoscedastic, meaning that all random variables have the same finite variance. If any of these assumptions do not hold, then ordinary least square estimates are biased and (or) inefficient. In a biased estimation, the expected value of a parameter estimate is different from the true value. However, if the bias shrinks as the sample size increases, the estimate is consistent. Inefficient means that an estimator is less accurate as the sample size increases than an alternate estimator.

In panel data analysis, we also need estimators to be consistent and efficient. Consistency in a panel data model means that the distribution of estimated coefficients, $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$, collapses on $\boldsymbol{\beta}$ as n becomes large: $\underset{n\to\infty}{\text{plim}}\,\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta}$. In other words, more observations lead to a more-accurate estimation. Also, the efficiency (minimum variance) of an estimator is a relative measure for a specific class of estimators. Where the variance of each estimator in a class is compared to the variance of all other estimators in the same class. For example, the pooled OLS estimator is efficient among the class of linear unbiased estimators.

Estimating pooled models resembles estimating a simple OLS model. In this case, a pooled panel model is considered as a long regression with $N \times T$ observations, and estimating with OLS method is as follows: $y_{it} = \alpha + X'_{it}\boldsymbol{\beta} + (\alpha_i - \alpha + \varepsilon_{it})$. The pooled OLS estimation is consistent when the independent variables are uncorrelated with the error terms. In contrast, if the true model is a fixed or random effects model, then the pooled OLS is inconsistent.

Estimating fixed effects models can take two main approaches: least squares dummy variable (LSDV) estimation and "Within" estimation. LSDV, which includes a large number of dummy variables in the model, is not an efficient approach when there are many individuals in a panel dataset. A "Within" estimator uses time-demanded dependent variables on the time-demeaned independent variables as in: $y_{it} - \bar{y}_i = (X_{it} - \bar{X}_i)'\boldsymbol{\beta} + (\epsilon_{it} - \bar{\varepsilon}_i)$. Relative to the general form in (3.3), the number of observations is still $N \times T$, but α_i, are removed. Using time-demeaned variables drops any time-invariant variables, such as α_i, that do not vary within an entity. Another technique to remove individual-specific effects from the model is through a a first-difference estimator, which uses one period change for each individual.

Different methods have been proposed for estimating random effects models, such as generalized least squares (GLS) and feasible generalized least squares (FGLS). The first method, GLS, estimates random effects models when a covariance structure of an individual is known; if a covariance structure is not known, FGLS works better, with a maximum likelihood estimation method and simulation (Baltagi and Chang, 1994). A random effects estimator uses a transformed model as follows:

$$y_{it} - \hat{\lambda}\bar{y}_i = (1 - \hat{\lambda})\mu + (X_{it} - \hat{\lambda}\bar{X}_i)'\boldsymbol{\beta} + v_{it} \tag{3.5}$$

where individual-specific effects are included in the error term as $v_{it} = (1-\hat{\lambda})\alpha_i + (\varepsilon_i t - \hat{\lambda}\bar{\varepsilon}_i)$, and λ is as follows:

$$\lambda = 1 - \sigma_u / \sqrt{\sigma_u^2 - \sigma_\alpha^2} \tag{3.6}$$

In equation (3.5) $\hat{\lambda} = 1$ and $\hat{\lambda} = 0$ correspond to a fixed effects estimator and pooled OLS estimator, respectively.

Model Selection and Validation

So far, we have discussed an individual-specific effects model, which is a one-way model (i.e., includes individual-specific effects). Additionally, if we include time-specific effects, δ_t, in a panel model, then we will have a two-way model as follows:

$$y_{it} = X_{it}'\boldsymbol{\beta} + \alpha_i + \delta_t + \varepsilon_{it} \tag{3.7}$$

The one-way model is prevalent in panel data research due to its parsimonious property, because the one-way model has less parameters than two-way models. Considering fixed vs. random effects, individual vs. time effects, and one-way vs. two-way effects provides different combinations of options for choosing a model.

To select an appropriate model in this research; first, we use the Breusch-Pagan (BP) Lagrangian Multiplier (LM) test (Breusch and Pagan, 1980) to determine whether our datasets are poolable or have individual-specific effects. The null hypothesis of the BP LM test says the variance across entities is zero, which means there is no individual-specific effect in the model. Rejecting the null hypothesis implies that the pooled OLS model is better (i.e., datasets are poolable).

If the the BP LM test reject their null hypotheses, then this indicates that the pooled OLS model is not appropriate, we need to compare the fixed effects model and the random effects model to choose the best one. For this purpose, we use the Hausman test, which has a null hypothesis of uncorrelation between an individual effect and independent variables (Hausman, 1978). If the Hausman test rejects the null hypothesis, then we select the fixed effects model as an appropriate model; otherwise, we would use the random effects model.

After choosing the best model, through panel data model-selection steps, we validate the estimated model by testing for autocorrelation and heteroscedasticity. The Breusch-Godfrey test investigates the existence of autocorrelation in panel models, and the Dickey-Fuller test (Dickey and Fuller, 1979) examines whether the series has a unit root (i.e., is non-stationary). To check for heteroscedasticity, which is the variance around the regression line is not constant for all predictor values, we use the Breusch and Pagan (1980) LM test. If autocorrelation and heteroscedasticity problems exist, then we use robust standard error correction methods to correct the standard error and calculate the adjusted p-values.

Finally, after validating the best model for each channel dataset, we compare the corrected estimates for both TC and MC channels to understand the behavior of the dependent variable (Channel Viewership), under two different IP protection policies – Track and Monetize. We determine which policy has the greatest (if any) effect on Channel Viewership, and also determine the magnitude of that effect.

3.5 Empirical Results

This section first presents model specifications and shows the result of estimating the MC and TC models, which use the Monetize and the Track policy, respectively. Then, it shows our diagnostic test results to confirm test validity. Finally, the section compares the most appropriate models for each channel to one another in order to assess the differences in the impact of UGC viewership on channel viewership.

The specifications of the pooled OLS, the fixed effects, and the random effects models we use are found in Table 3.2. The independent variable in all models is CV_{it} and the dependent variable in all models is UV_{it}. CV_{it} is the Channel Viewership, and UV_{it} is UGC Viewership, both are defined for time (day), t, and UGC video i. In particular, for a given t, CV_{it} is the same, as it is the daily channel viewership, while UV_{it} corresponds

Table 3.2: Panel models specification

Model name	Specification
Pooled OLS	$\ln(CV_{it}) = \alpha + \beta \ln(UV_{it}) + \varepsilon_{it}$
Fixed effects	$\ln(CV_{it}) = \alpha_i + \beta \ln(UV_{it}) + \varepsilon_{it}$
Random effects	$\ln(CV_{it}) = \alpha + \beta \ln(UV_{it}) + (\alpha_i + \varepsilon_{it})$

to the number of daily views of UGC video i. Constant α is the model intercept, $\boldsymbol{\beta}$ is the coefficient of the independent variable, α_i is individual effects, and ε_{it} is the error term. Table 3.3 shows the regression results for the pooled OLS, fixed effects, and random effects models for MC. The analog for TC is in Table 3.5.

3.5.1 Monetizer Channel Model Estimation

Table 3.3 shows the estimation of the independent variable, UGC Viewership (UV), and coefficient, along with the standard error in parenthesis for the Monetizer Channel (MC). P-values for the pooled OLS, the fixed effects, and the random effects coefficients are 0.2189, 0.2885, and 3.543e$-$05, respectively. Thus, based on the p-values, the reported coefficients of the pooled OLS model, the fixed effects model and the random effects model are statistically significant.

Table 3.4 shows the results of the LM test and the Hausman test that help us to identify the best MC model. Based on the first row of Table 3.4, the p-value of the LM-test, which helps to identify individual effects, is < 0.001. This p-value leads us to reject the null hypothesis, and we are statistically confident that the pooled OLS model is not better than the random effects model.

So far, the LM test result confirms that the pooled OLS model is not appropriate. Thus, we exclude the pooled OLS model and use the Hausman test, to determine if the fixed or random effects models are most appropriate for our data. The Hausman test

Table 3.3: Panel data estimation - MC

	Dependent variable:		
	Channel Viewership		
	Pooled OLS	Fixed Effects	Random Effects
UGC viewership	0.002***	0.002***	0.001**
	(0.0001)	(0.0001)	(0.0003)
Constant	−0.00004		−0.0001
	(0.0001)		(0.0001)
Observations	968,204	968,204	259,892
R^2	0.0004	0.0004	
Adjusted R^2	0.0004	0.0003	
F Statistic	341.995***	335.541***	
	(df = 1; 968202)	(df = 1; 948664)	
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 3.4: Monetizer channel model selection results

	Statistics (Chisq)	p-value
Breusch-Pagan Lagrange Multiplier Test	1074.16.2	<0.001
Hausman Test (Fixed vs. Random)	36.12	<0.001

is used to determine if the error term, ε_i, is correlated with the independent variables. The Hausman test's null hypothesis is that the error terms are uncorrelated with the independent variables. The Hausman test results show that the χ^2 statistic is 36.12, leading to a p-value near zero. Consequently, we can reject the null hypothesis and conclude that the random effects model is better suited for our data than the fixed effects model for the MC panel dataset.

3.5.2 Tracker Channel Model Estimation

Table 3.5 compares the results obtained from the fitting the panel data models on the TC panel data.

In the random effects model, the coefficient of the independent variable is significant.

Table 3.5: Panel data estimation - TC

	Dependent variable:		
	Channel Viewership		
	Pooled OLS	Fixed Effects	Random Effects
UGC viewership	0.003	0.002	0.015***
	(0.002)	(0.002)	(0.004)
Constant	-0.007^{***}		-0.007^{***}
	(0.001)		(0.001)
Observations	141,286	141,286	42,834
R^2	0.00001	0.00001	
Adjusted R^2	0.00001	0.00001	
F Statistic	1.512	1.127	
	(df = 1; 141284)	(df = 1; 137911)	
Note:			*p<0.1; **p<0.05; ***p<0.01

As with the MC models, we use LM and Huasman testing to select the best TC model.

Based on Table 3.6, the p-value of the LM test is near zero. As a result, we conclude that the pooled OLS model is not appropriate since the LM test detects endogeneity in the TC dataset, suggesting that we need to consider individual effects in the panel dataset. The Hausman test results show the χ^2 test statistic is 20.86, which offers a p-value near

Table 3.6: Tracker channel model selection test results

	Statistics (Chisq)	p-value
Breusch-Pagan Lagrange Multiplier Test	181.24	<0.001
Hausman Test (Fixed vs. Random)	20.86	<0.001

zero. As such, we reject the null hypothesis stating that the Random Effects model is better than the fixed effects model, and accept the fixed effects model as the most appropriate model for the TC panel dataset.

As an alternative to the Hausman test, Mundlak (1978) offers a new approach for selecting between fixed effect and random effects models when the error terms are heteroskedastic. Mundlak testing is based on a random-effects model where the individual means of the time-varying characteristics are added to the model. Then, an F-test on the coefficients of the individual means of the time-varying characteristics helps determine whether to use a random effect model or fixed-effect model in our analysis. Rejecting the Mundlak test's null hypothesis (i.e., coefficients of means of the time-varying independent variables are jointly zero) means that the fixed-effect model is better than the random effect model.

To cross-check our model selection process we use Mundlak testing for both MC and TC datasets. The Mundlak test statistics (χ^2) for MC and TC models are 4.1 and 53.5, respectively, with p-values of less than 5%. Then, we reject the null hypothesis of the Mundlak test and confirm that the fixed effect models are better for both datasets.

3.5.3 Model Estimations Comparison

Tables 3.7 compares the MC and TC fixed effects models. The F Statistics for MC model is significant, with a p-value < 0.01. The F test is used to determine if all coefficients in a model are not zero. The small p-value of the F test implies that the MC model has at least one coefficient statistically different from zero. However, the p-value for the TC model suggests we cannot statistically differentiate any estimated coefficient from zero for that model.

3.5.4 Robust Standard Error Analysis

The best-fit models, fixed effects model for both MC and TC assume certain conditions to be true. We now validate these assumptions, to ensure we may indeed apply the fixed effects models to our data.

For the model validation procedure, we start with the Augmented Dickey-Fuller (ADF) test to check for stochastic trends in our datasets. The null hypothesis is that the series

Table 3.7: Fixed effects estimation comparison - TC and MC models

	Dependent variable:	
	Channel Viewership (CV)	
	Monetizer Channel fixed effects	Tracker Channel fixed effects
UGC viewership (UV)	0.00235315***	0.0024536
	(0.00012846)	(0.0023114)
Observations	968,204	141,286
R^2	0.0004	0.00001
Adjusted R^2	0.0003	0.00001
F Statistic	335.541***	1.127
	(df = 1; 948664)	(df = 1; 137911)
Note:		*p<0.1; **p<0.05; ***p<0.01

is non-stationary or has a unit root. The ADF test for both TC and MC datasets shows that there is no unit root problem, and we can reject the null hypothesis of ADF with p-value 0.01 for both models. We also use the Central Limit Theorem (CLT), to argue that residuals in our tests will be normally distributed as we have over 100,000 observations in each dataset.

Next, we need to make sure our models do not violate the assumptions of no autocorrelation in error terms and constancy of errors' variance (i.e., homoscedasticity). For testing the presence of autocorrelation in panel models, we use the Breusch-Godfrey/Wooldridge test, which rejects its null hypothesis and shows that autocorrelation remains in the idiosyncratic error terms (Breusch, 1978; Godfrey, 1978; Wooldridge, 2010). The p-value of both tests are near zero, and the chi-squared test statistics are 56892.97 and 334557.6 for TC and MC models, respectively.

Subsequently, to identify heteroscedasticity, which is the counterpart of homoscedasticity, the student Breusch-Pagan tests show that both MC and TC models suffer from heteroscedasticity. Meaning that the error terms are randomly scattered around the regression line, but the variance is not constant over time. Breusch-Pagan test statistics for the TC and MC models are 21.71 and 82.40, respectively, the p-values near zero.

Finally, considering the test results in the second and third steps, we conclude that our models have autocorrelation and heteroscedasticity problems. Thus, the estimated standard errors and the reported p-values are not correct, and they need to be adjusted.

The literature offers two main approaches to correct the bias in the estimated standard errors, heteroscedasticity and autocorrelation consistent (HAC) estimators or feasible generalized least squares (FGLS) estimators (Reed and Ye, 2011). The econometric literature offers different procedures for heteroscedasticity and autocorrelation consistent (HAC) covariance estimation (White et al., 1980; MacKinnon and White, 1985; Newey and West, 1986). In this research, we use HAC estimators to account for heteroscedasticity and autocorrelation simultaneously while generating panel data standard error estimators (Musau et al., 2015).

White et al. (1980) offer HC_0 as an initial HAC estimator, but MacKinnon and White (1985) raised concerns about the performance of HC_0 in small samples, and introduces HC_1, HC_2, and HC_3 to address this issue. In addition to this, Stock and Watson (2008) show that the White et al. (1980) robust errors are inconsistent in the case of the panel fixed effects regression model. Thus, we use the HC_3 estimator, which is a better estimator than HC_0 based on a Monte Carlo simulation (Scott Long and Ervin, 1998), to address both heteroskedasticity and autocorrelation at the same time in fixed effects models.

Table 3.8 shows the TC and MC fixed effects models with robust standard errors using the HC_3 estimator. Comparing Table 3.8 to Table 3.7, the coefficients of our independent variable, UGC viewership (UV), for both models are the same, however the standard errors and corresponding p-values are not.

HC_3 is one method found in the literature to address heteroskedasticity. Another method is the cluster-robust variance estimation (CRVE) (Wooldridge, 2003; Petersen, 2009; Cameron et al., 2011). We cross-check the results in Table 3.8 with the CRVE approach. The failure to control for clustered correlated errors leads us to overstate t-

Table 3.8: TC and MC models with robust standard error estimation

Robust Variance Estimation	Estimate	Std. Error	t value	p-value
UGC viewership (UV) (TC model)	0.0024536	0.0027337	0.8975	0.3694
UGC viewership (UV) (MC model)	0.00235315	0.00015726	14.964	$<$ 2.2e-16

statistics and understate of p-values, that might provides too narrow confidence intervals. We use the Satterthwaite (1946) approximation for t-tests proposed by Bell and McCaffrey (2002) for applying the CRVE on our panel dataset. Table 3.9 shows clustered standard errors estimation using the Satterthwaite (1946) approximation.

Table 3.9: TC and MC models with cluster-robust standard error estimation

Cluster-Robust Variance Estimation	Estimate	Std. Error	d.f.	p-value
UGC viewership (UV) (TC model)	0.0024536	0.00273	1674	0.37
UGC viewership (UV) (MC model)	0.00235315	0.000157	8311	$<$ 0.001

Comparing Tables 3.7, 3.8 and 3.9, we note that the coefficients of UGC viewership (UV) are the same. The standard errors in Tables 3.8 and 3.9 are corrected and the p-values are different from those in Table 3.7, but the robust estimation of TC model coefficient is still not significant and the MC model is significant. As such, we conclude that the results of initial fixed effects estimations are valid in even in case of autocorrelation and heteroskedasticity.

To ensure the reliability of our analysis, we have a large sample size in both the TC and MC models (see Table 3.1). However, the extremely large samples may cause problems in our statistical inferences and result in a very low p-value Lin et al. (2011). To make sure our large sample sizes do not affect our inferences, we randomly sub-sampled both datasets 100 times with different sample sizes for our statistical tests. The results of these sub-sampled tests show p-values of the same significance with the smaller samples (see Appendix A.1 for more details).

3.6 Conclusions

The primary objective of this research is to clarify for YouTube partners and content owners: what is the impact of UGC video viewership on the YouTube channel viewership under two different YouTube IP protection policies, where the YouTube channel holds the copyright of the UGC content?

In this research, we studied the impact of two different policies, Monetize and Track, on YouTube channels viewership, using a panel data analysis. The findings are summarized in following main points:

- We considered three model specifications (i.e., pooled OLS, the fixed effects, and the random effects model) in our panel data analysis. Using appropriate tests we conclude that the fixed effects model is most appropriate for the TC and the MC datasets (see Sections 3.5.2 and 3.5.1).

- After choosing the fixed effects model as an appropriate model specification, we validate the estimated fixed effects models by testing for autocorrelation and heteroscedasticity. We account for autocorrelation and heteroscedasticity by adjusting the standard errors, using the appropriate methods (see Section 3.5.4).

- The TC model's dependent variable (CV) coefficient is 0.0024536. The coefficient of TC model is not statistically significant with a p-value of 0.288. Thus, we conclude that UGC Viewership related to the channel with the Track policy does not significantly affect Channel Viewership (reject H_1^{TC}).

- The MC model's dependent variable (CV) coefficient is 0.00235315. The coefficient of MC model is statistically significant with a p-value of < 0.01. Consequently, we conclude that viewership of UGC videos infringing on the MC channel IP does significantly affect channel viewership in a positive manner (accept H_1^{MC}).

Thus, one percent view of UGC affects MC channel viewership by about +0.0024 percent under the Monetize policy.

The results of the TC model provide additional insights on the impact of a copyright holder making his/her intellectual property openly available (e.g., Tesla open-source pledge to make all its patents freely available) and how this decision influences the copyright holders' business (Musk, 2019). However, the TC model results suggest that copyright-free music does not attract more viewers to come to the original channel.

This research, however, is subject to certain limitations. First, we did not conduct a controlled-experiment; therefore, we compared two different channels to evaluate the IP protection policies' available on YouTube. In particular, it is not clear if there are endogenous properties of each YouTube channel that lead to the observed results, or if the results are due to the different IP protection policies employed by each channel. The second limitation concerns the generalisability of these results to all YouTube channels. Both channels we consider only have music videos, that are not visual in nature, i.e., a "viewer" may turn on the videos but never watch. Not having to watch the videos we consider may mean that our results only hold for videos that do not require active viewers. Despite these limitation, the study certainly can be generalized to YouTube music channels, which is one of the top four content categories watched by YouTube users (Google, 2016*d*). Our results are a needed first step in helping a YouTube channel determine what IP protection policy is best suited for its videos. The steps we take in our analysis may be carried out by interested channels to compare different policies across their own videos to determine how each policy impacts metrics of interest.

This research's datasets are different in the number of observations; however, each dataset is large enough to have statistically significant results (see Table 3.1). Nevertheless, we subset the MC dataset (1.1 million observations) randomly to generate a subset that is the size of the TC dataset (155 thousand observations) to make sure that the size of the

MC datasets does not affect the model estimations and its statistical significance. Then, we tested the subsetted MC datasets, and the results are the same, which means that the number of observations in the MC dataset does not affect the significance of the estimated coefficients.

The only independent variable is UGCs' viewership in our models' specifications (see Table 3.2). However, we also have access to the other independent variables, such as the number of likes, dislikes, and comments for each UGC, but we did not include them in our models' specifications for two reasons. First, these variables have a lot of zeros. Second, adding these variables did not change the significance of UGC viewership and constant values in both TC and MC models.

Our empirical analysis takes a first step towards analyzing whether the most-common IP protection method used on YouTube, content owners choose to monetize user-generated content, is in the best interest of a IP owner? The significant coefficient of the MC model means that using copyright protection policies does not only help to earn money through monetizing UGCs but also increases channel viewership.

Chapter 4

A data-driven statistical framework for enhancing business revenue generating processes: An application to advertising policies on YouTube

4.1 Introduction

Internet advertising revenues in the United States was more than $107 billion for the year 2018 that was increased by 21.8% over 2017 (AIB, 2018). YouTube content creators (companies) generated 2 billion US dollars in 2016 (Hamedy, 2016) by showing different types of ads on posted videos. The money generated by YouTube is attracting more channels as there is a 50 millions creators on YouTube in 2019 (Omnicore, 2019). However, YouTube channel owner must decide on what type of advertisements to show on their posted videos. Currently there is no clear guideline to help a channel owner (company) determine the profit maximizing set of ads to show on their videos. With this research, we develop a statistical framework that YouTube companies can use to maximize their profits.

YouTube does not allow companies to pick which ad to show. However, YouTube allows companies to pick what types of ads to show on each video posted, please see Table 4.2 for more details. With our method, companies will be able to determine revenue-maximizing advertisements by means of a statistically valid approach. We run our developed statistical framework via an experimental study to answer an essential question for YouTube companies: what types of ads should a YouTube company show on posted videos? Even a 1% increase in revenue will result in millions of additional US dollars for large companies. The remainder of this section first discusses the importance of online advertising and its financial value. Next, we briefly introduce online advertising platforms, such as the YouTube monetization program, and finally, we review the paper's structure.

4.1.1 Online advertising

In 2017, Canadian internet advertising revenue was near $6.8 billion, up 23% from 2016; and 2018 revenue was forecast to increase 14%, to hit $7.7 billion (IABCanada, 2018). In fact, most web firms now generate their income through advertising; for example, in 2017, Google earned $36.7 billion, which is 77.8% of US search ad revenues (Marvin, 2017).

According to the *Cisco Visual Networking Index report*, Internet video traffic will grow four-fold from 2017 to 2022, and web users will increase from 44% of the world population to 58% between 2016 and 2021. Video accounted for 73% of internet traffic in 2016 and will make up 82% of all internet traffic in 2021. People are continually consuming more internet traffic because of the better quality video, and it is estimated that by 2021, every second, a million minutes of video content will cross the network (Cisco, 2017).

Based on Sandvine's *Global Internet Phenomena Report*, YouTube is one of the top websites, and responsible for consuming the most network bandwidth in most regions. In North America, Netflix and YouTube are two top applications and use up 55% of all downstream internet traffic. In 2016, YouTube alone accounted for 19.16% of peak downstream traffic (Sandvine, 2016). This giant video-sharing website, has over a billion users (which is nearly one-third of all global internet users) who watch billions of hours of video that generate billions of views per day. More than half of the total YouTube views come from users' mobile devices. YouTube enables companies to upload and share their videos with others. Companies may also earn money through uploaded videos by having viewers view advertisements before, during, or after the creator's video. For example, as of March 2015, more than $100,000$ videos created at YouTube Spaces had generated over 1 billion views, and the number of YouTube channels earning six figures per year is up 50% each year (YouTube, 2017*a*).

Video advertisements are a strategic and innovative tactic for attracting audiences because they communicate with customers to tell a brand story and explain companies' value propositions (Kallas, 2018). Based on the *Ipsos Connect Report*, almost 50% of Web users look for videos related to a product before visiting a store (Google, 2016*c*).[1] Moreover, Web users who view product videos are 1.81 times more likely to purchase than non-viewers (Adobe, 2015). Furthermore, based on HubSpot's *State of Inbound 2017 Report*, 48% of marketing teams plan to add YouTube to their marketing efforts in the next 12 months (HubSpot, 2017), since brands that take advantage of video advertising grow their revenue 49% faster than non-video advertising brands (vidYard, 2014).

Among video-sharing websites, YouTube is the most popular and accounts for two-thirds of premium online video-watching across all devices (Google, 2016*a*). YouTube claims that eight out of ten, 18-to-49-year-olds watch YouTube in an average month (Google, 2016*b*). Moreover, it is important to know that people are changing the way they interact with content; 55% of individuals consume video content thoroughly (HubSpot, 2016).

[1] Google / Ipsos Connect, March 2016, GPS Omnibus, n=2,013 US online respondents 18+)

4.1.2 YouTube monetization

The YouTube advertising ecosystem consists of three principal entities: viewers, creators, and advertisers. Viewers, who are the most important entity, come to YouTube channels to listen, watch, and learn. Viewers' engagement with YouTube's videos are at the core of creators, advertisers, and YouTube itself making money. Creators are companies who share their knowledge and creativity for different purposes, such as making money from their channel audiences. Advertisers try to find appropriate YouTube channel categories that will connect advertisers with their target audience.

YouTube monetization programs connect viewers, creators, and advertisers and attempt to create value for all three entities. The AdSense program places ads, provided by advertisers, inside or near videos whose creators have enabled monetization. To generate money from videos, creators must have worldwide commercial rights to everything in the video. When content creators choose to show advertisements, they may select the type of advertisement to show, for example, video (desktop and mobile advertising), text and banner (display advertising), image and pop-up (social media advertising), etc (TubeMogul, 2017).

Prior to our work, it is unknown what advertising format maximizes a firm's revenue. More importantly, to our knowledge there is not framework a company may use to determine what advertising formats maximize revenue. This study investigates which ad formats are used across companies videos, and compares how revenue and viewership are affected by the different ad formats.

The remaining part of the paper proceeds as follows: Section 4.2 gives a brief overview of the related literature and presents a formal problem statement. Section 4.3 describes the methodology used for this study and the general framework used which is a type of experimental design. Section 4.4 presents the application of the framework to a real-world YouTube company; and Section 4.5 concludes the paper and presents future research directions.

4.2 Literature Review and Problem Statement

This research relates to three streams of literature. One major stream and two minor streams. The major stream of literature is Revenue Management (RM) models. The two minor streams are 1) multi-armed bandit and 2) assortment planning problems. We now address each stream, major and minor in turn. Conclude the section by formally stating our research problem.

The RM models for the online advertising tend to take an industry wide lens, where limited capacity for advertisements is matched to a potentially limited set of candidate advertisements. Next we review related studies in RM applied to online advertising.

The initial interest in RM models arise with the desire for capacity management in the aviation and hotel industries to manage overbooking (Belobaba, 1987, 1989; Rothstein, 1971, 1974; Smith et al., 1992; Littlewood, 1972). These initial studies assume fixed prices in their analysis. However, in the real world problem, other pricing strategies are used to solve business problems. For example, a new stream of studies emerged to deal with different pricing methods, and dynamic pricing became an integral part of RM models (Feng and Gallego, 2000; Bitran and Mondschein, 1997; Feng and Gallego, 1995; Gallego and Van Ryzin, 1994).

Dynamic pricing strategies help industries that face uncertain demand (i.e., stochastic and price-sensitive) and unstable supply (i.e., perishable capacity and selling-time limits) to balance profitability and utilization of their available capacity (Bitran and Caldentey, 2003). Many studies discuss the application of dynamic pricing strategies in the RM model in different industries, such as retailers, car rental agencies, hotels, Internet service providers, airlines, and passenger railways (Subrahmanyan and Shoemaker, 1996; Bitran and Gilbert, 1996; Carroll and Grimes, 1995; Ciancimino et al., 1999; Geraghty and Johnson, 1997; Nair and Bapna, 2001). A book by Talluri and Van Ryzin (2004) is a comprehensive reference of classic revenue management models. However, this book does not cover revenue management problems for an online environment, such as demand and supply uncertainty over a specific time horizon.

Araman and Fridgeirsdottir (2011) study uncertainty of supply and demand in online advertising. They use dynamic pricing models to maximize a web publisher's profit while satisfying advertisers' campaign requests. Their research uses the cost-per-impression (CPM) model (i.e., an advertiser pays every time an ad is displayed) in their analysis. Araman and Fridgeirsdottir provide a capacity allocation mechanism that shares the capacity among different advertisers, and manages demand uncertainty for display slots while considering an uncertain supply (i.e., viewership).

The cost-per-click (CPC) model is the other online advertising pricing model commonly used by researchers. For example, Fridgeirsdottir and Najafi-Asadolahi (2009) uses the CPC model when ad requests come from an advertising network. Chen et al. (2014) discuss the real-time-bidding (RTB) model to propose a dynamic mathematical pricing model that helps allocate and price future impressions in real-time auctions for display advertising. In addition to pricing challenges, content owners/potential sites that show advertisements must find a balance between ad revenue and subscription fees. For example, Kumar and Sethi (2009) provide a dynamic pricing strategy to maximize revenue by resizing advertisement spaces and adjusting subscription fees. Yang et al. (2010); Chickering and Heckerman (2003); Radovanovic and Zeevi (2010) all discuss the challenges of managing advertisement inventory and its dynamic allocation. However, none of these papers discuss what ad formats are most appropriate for a company to maximize online advertisement revenue, a gap we address in this research.

We now move to the other two minor streams of literature, multi-armed bandit and assortment planning problems in the online advertising industry. A multi-armed bandit (MAB) problem is associated with RM, and deals with the allocation of limited resources among competing choices to maximize revenue over time (see Lai and Robbins (1985) and Auer et al. (2002) for the MAB classic formulation and solution). In digital marketing research, a MAB is used to allocate a advertisements to advertising spaces over time so as to maximize revenue. For example, Abe (1999) uses a dynamic bandit formulation to manage display ads scheduling in online advertising. Furthermore, Pandey et al. (2007) use the MAB formulation to find relevant ads to display in specific web pages; and Agarwal et al. (2009) use the same formulation to find ads that optimize click-through rates (CRT).

Combining of different ad formats to maximize advertiser revenue is similar to the notion of assortments in dynamic assortment planning problems in the retail industry. Choosing the assortments that maximize expected revenue appears in the RM literature (Bernstein et al., 2015; Sauré and Zeevi, 2013). In the online advertising industry, ad publishers have limited access to users' profiles and assume that users are homogeneous in their preferences. Advertisers must decide dynamically which set of products to display to users, similar to what advertisers must do when deciding which ads to show to a user.

To the best of our knowledge, all the studies above consider different pricing strategies in revenue management models. They also disregard a user's profile and preferences when customizing their display decisions. In contrast, our research focuses on a problem in which an advertiser chooses the specific ad-format settings that will maximize revenue. As such, we have not capacity limits, but in our model we do not know how users will collectively respond to different types of advertisements.

4.2.1 Research problem

Advertisers, creators (companies) and viewers have decisions to make when using YouTube. Advertisers must determine their ad type (see Table 4.2), ad content (what is in the ad), and ad audience (which YouTube viewers will see the ad). companies determine video content, genre, length, so as to attract an audience and generate revenue. In addition, companies may choose the ad formats and positions for each video posted. Finally, viewers decide which videos to watch and how to engage with creator and advertiser content.

One of the main questions that companies want to answers is: what set of ad formats will maximize channel revenue? This research partially addresses the question above. In particular, we help creators determine what ad formats to use in order to optimize two key performance indicators (KPIs) of interest across their entire channel. The two KPIs we consider are *average views* and *average gross revenue*. These two KPIs are determined through a discussion with an industry collaborator. We consider KPI performance for the entire channel and not an individual video, as the entire channel is not as biased by the natural KPI day-to-day fluctuations that occurs on YouTube. The research questions of this paper are as follows:

- What is the difference for YouTube channel **average viewership** between the various combinations of advertisement formats run on companies' videos?
- What is the difference in YouTube channel **average gross revenue** between the various combinations of advertisement formats run on companies' videos?

Currently, most YouTube companies try to optimize profits by selecting some or all forms of advertisement for their videos based on intuition or trial and error[2].

To answer these questions, this paper first presents the framework we developed to determine what advertising formats maximize a company's revenue. Second, the paper presents an application of our framework in the form of an experimental study conducted on a set of YouTube videos owned by an industry partner. The results of the application

[2]YouTube even informs content-owners that non-skippable ads may decrease viewership.

inform our industry partner on the set of ad formats that maximize a channel's viewership and revenue.

4.3 Data-Driven Statistical Framework

In this section, we briefly present one of the main contribution of our research, a data-driven statistical framework (DDSF), to optimize companies' KPI of interest. The proposed DDSF is a comprehensive tool that applies to various online advertising platforms and KPIs. This framework enables companies to measure and optimize various performance metrics via a robust experimental design.

Figure 4.1 shows the main components of our DDSF and their connections that are categorized in four stages 1) Design of Experiment, 2) Data Collection, 3) Running Experiment, and 4) Making Decision (dashed arrows are optional steps). In the remainder of this section, we describe each category briefly, and then we provide more details on 3 and 4.

- **Design of Experiment**: This first stage determines the framework's settings that consist of three components as follows:
  - First, we select the **subjects**, in our case it is a media platform on which we want to implement the DDSF. There are a growing number of various platforms that advertise to their online users, such as video-sharing websites, social media (e.g., Facebook and Instagram), email marketing campaigns, web pages (e.g., blog and online magazines), mobile apps, search engines (e.g., Google ads), etc. For example, in our case study, we select a YouTube music channel as a platform that displays various advertisement formats (i.e., subjects are YouTube music videos).

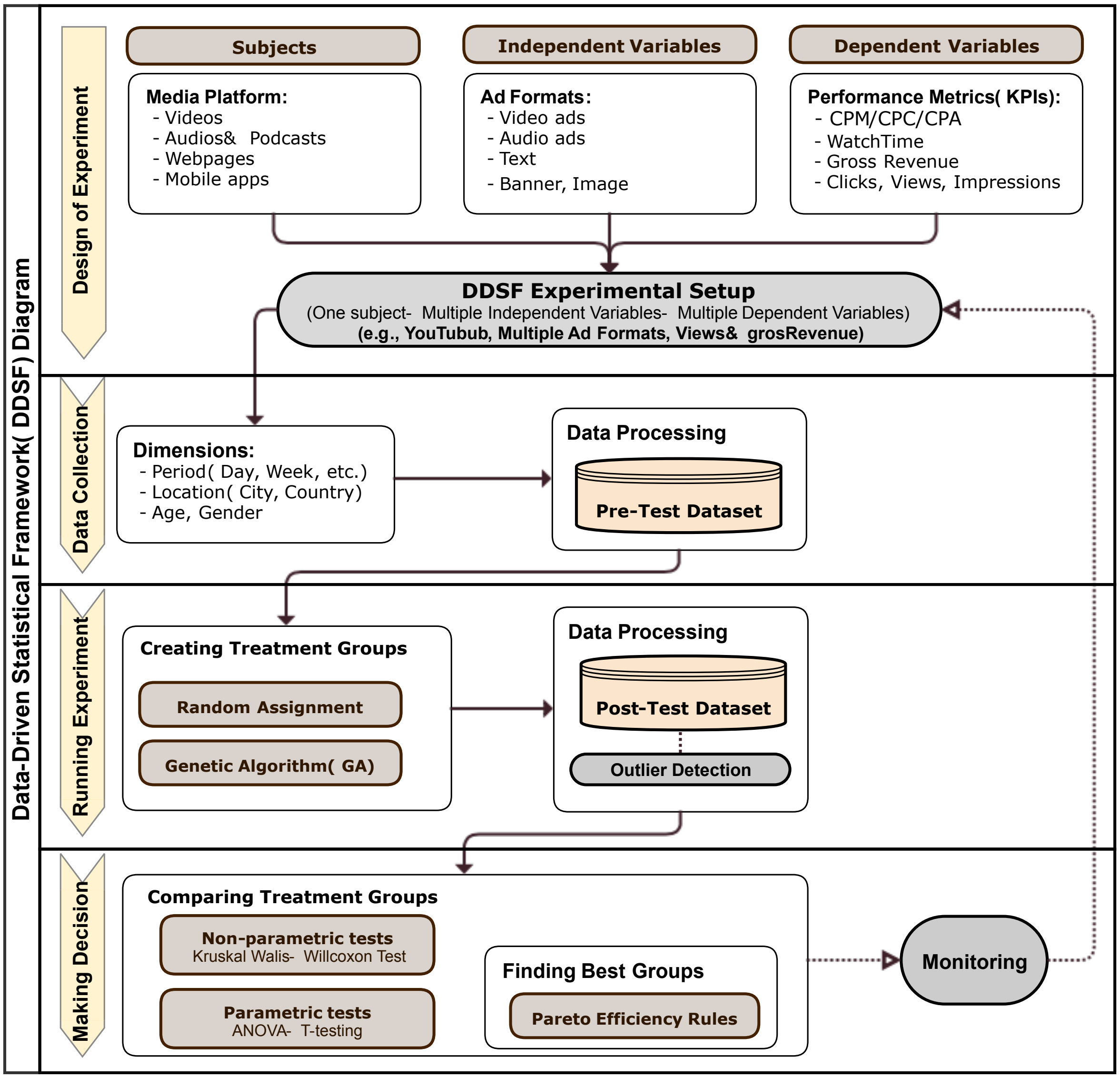


Figure 4.1: Data-Driven Statistical Framework Diagram

– Second, we select our **independent variables** (a.k.a. decision factors, attributes, explanatory variables) that are a set of variables that we may change that their values may influence the dependent variables. In our case study, the independent variables are ad formats. Each platform potentially has its own types of ad formats. For example, an online magazine can show image banners, text box, or video ad in its web content. YouTube shows five different ad formats on its platform (see Table 4.2).

  - Third, we select the **dependent variables** that are KPIs available from the platform on which the DDSF will run. The DDSF will determine what values of the independent variables result in the statistically best values of the dependent variables. To summarize, in our case study the Subjects is a YouTube music channel, the independent variables are ad formats, and the dependent variables are the YouTube music channels average views and average gross revenue. In general, the dependent variables are a function of the selected platform and potentially selected independent variables. For example, in a video-sharing platform "WatchTime" is a meaningful metric for a video ad, and it is not valid in a banner display platforms.

- **Data Collection**: Prior to running an experiment we must first establish a baseline. We do this by collecting the **pre-test dataset**. Collecting data has the following two components:

  - **Dimensions** are dataset's attributes that enable users to retrieve different granularity (i.e., the level of detail at which data are stored in a database) of datasets. For example, time period (e.g., hour, day, or week) or a specific user's demographics (e.g., age, location, or gender) are different dimensions. Depending on the selected platform, from the stage above, the set of available dimensions is determined. It is also a business decision as to which dimensions are important to the company. Using the selected dimensions we derive the KPIs and other experiment variables.

  - A **database** to store the collected pre-test datasets, using the dimensions above.

- **Running Experiment**: In this stage, we run the experiment. Running the experiment requires the following three components:

  - **Creating treatment groups**: first, we analyze the pre-test datasets to form groups of subjects to assign treatments to each group. Treatments are determined by the independent variables in the Design of Experiment stage (see Section 4.3.1 for more details on the design of the experiment). Creating homogeneous decision groups is an essential stage in our DDSF. Groups are said to be homogeneous if there is not statistical difference between the groups on the KPIs of interest prior to treatment. To make homogeneous groups, we first use random assignments, and then we use statistical tests to assure that created groups are homogeneous. If the random assignment procedures nonhomogeneous groups, then we use a genetic algorithm to create homogeneous groups (see Algorithm 1). After creating homogeneous groups, we assign a treatment to each group.

  - **Post-test dataset collection**: after treating each group, we collect data post-treatment. The collected data is one required to determine the KPIs of interested selected in the Design of Experiment stage.

    * Determine if the post-test datasets are normally distributed.

    * **Data processing (remove outliers)**: We use two methods for detecting outliers. The first method only uses the collected post-test dataset. Using the Standard Boxplot when we have a normal distribution, or Adjusted Boxplot when dataset distribution is not normal, we identify and remove outliers. The second method considers any unusual behavior or changes in the media platform's traffic sources (see case study in Section 4.4 for more details).

- **Making Decision**: After running the experiment, we make decision to select best treatment groups. Making decisions requires the following three components:

- **Comparing decision groups**: If datasets are normally distributed, then we use parametric tests (e.g., t-test or ANOVA) to compare the decision group's performance. If datadets violate the normality assumption, then we use non-parametric tests (e.g., Kruskal Wallis test and Wilcoxon post-hoc test) to compare decision groups (please see Section 4.3.2 for more details). After comparing all treatment groups to one another, we provide insights to the channel owner as to what treatments perform best. It is then up to the channel owner to determine what changes she like to make on the types of ads to run on her channel. If we have selected one metric (in the first stage), then we can compare the average value of metrics and find the best treatment. In the case of two or more metrics, we use the **Pareto efficiency rules** to find the treatments that are Pareto dominant.
- Finally, we may choose to run the experiment again with a smaller set of treatments to further statistically compare and **monitor** the efficacy of the selected treatments.

In the following sections, we provide more details on design of experiment and non-parametric statistical tools proposed on the DDSF.

4.3.1 Design of experiment

Design of experiment (DOE) is a method used to develop a framework one may use to establish a cause-and-effect relationship between independent and dependent variables using guided trials. DOE has two main classes: randomized experiments and quasi-experiments. Randomized experiments include control or comparison groups and uniformly random assigns non-control subjects to treatment groups. The quasi-experimental design lacks one of the two critical features of randomized experiments (i.e., comparison groups or random assignments to subjects to treatments is not done uniformly at random).

A uniformly random assignment is a common practice that lessens experimental bias by balancing the inclusion or exclusion of other unobserved variables in treatment groups. However, uniformly randomized experiments may not be the best experiment type to run, as each subgroup needs to be homogeneous (Vargas et al., 2017). Particularly, if the distribution of participant measurements is non-uniform, then there probablistically uniformly allocated subjects to treatments may result in non-homogeneous treatment groups. Berk et al. (1988) discusses situations in which uniformly random assignment may fail in the implementation of field experiments, and provides four actions to mitigate the risk of non-homogeneous treatment groups. Quasi-experiment with a control group is very much like a randomized experiment except that subjects are assigned to treatments in non-uniformly random fashion, to ensure treatment group homogeneity (Campbell and Cook, 1979).

In our study, to ensure treatment homogeneity, we develop a quasi-experiment. We utilize a genetic algorithm (GA) to assign subjects to treatment groups, and the GA improves, relative to uniformly random assignment, the homogeneity treatment group prior to treatment.

Finding an exact solution for assigning subjects to treatment groups is a large-scale combinatorial optimization problem, with approximately 110 videos per group and 8 groups, there are 880 choose 8 possible assignments, which is more than the estimated number of stars in the universe. The large number of possible solutions–too many for a computer to search for all. For instance, first, we tried to use used mixed-integer programming (MIP) to find an accurate solution for treatment assignment in which all groups are similar. However, the integer variables in our MIP model made the optimization problem non-convex, resulting in a difficult and time-consuming problem to solve. To overcome this computational issue, we turned to heuristic approaches to find a good solution. We have no idea what is optimal and we did not check duality gaps to make any claims about optimality or near optimality.

In our treatment assignment problem, we found not tractable exact algorithm. Thus, we initially try to use a randomized search algorithm to narrow down the search set and find a solution. Any search algorithm that solves our treatment assignment problem will work. However, this research uses GAs because they are practical and quick to implement. Additionally, a GA requires no information about the structure of the objective function and explores the search space, by design. Additionally, GA uses both crossover and mutation operators, which makes it less susceptible, relative to other heuristics, to terminate at a local optima. We remind the reader that there are other methods one may use. However, as the purpose of this research is to determine the type of advertisements to run on YouTube videos, once an acceptable group allocation method is found, we did not dwell on other methods.

Zabinsky (2010) categorizes the GA as a type of population-based random search algorithm that uses randomness to create an initial population and new individuals via mutation and crossover operations. As the first step in the GA, we randomly initialize populations. For a given population, the GA determines the best items of the population using a fitness function to inform the next population. Then, until convergence, the GA iterates the following four steps: 1) select parents from the population (using the fitness function), 2) generate a new population with crossovers (mixes selected parents), 3) mutate the new population (introduce slight variations in the new population), and 4) calculate fitness scores for the new population (see Algorithm 1 for the GA implementation algorithm).

Our proposed framework uses a post-test quasi-experimental design that uses GA treatment assignment as an alternative to random assignment. The general notational form of our experimental design is as follows:

$$X \qquad O_{i,t+1} \;\; O_{i,t+2} \;\; O_{i,t+3} \;\; \ldots \;\; O_{i,t+n} \tag{4.1}$$

where X is a treatment group (i.e, ad formats), t is a time index, i is a KPI index, $O_{i,t+1}$ is the value of KPI number i (i.e., *views* or *gross revenue*) at time t, and n is the number of observations. The next section describes the non-parametric statistical tools proposed by our DDSF to compare treatment groups.

4.3.2 Comparing treatment groups

The primary statistical tools used to analyze any experimental designs are parametric ones, such as significance testing (t-test), analysis of variance (ANOVA), regression, correlation, and residual analysis. All these methods focus on the population's parameters or probability distributions that depend on specific assumptions. For example, ANOVA, which is a standard method used to compare two or more treatment groups, has three main assumptions: 1) the dependent variable distributions are normally distributed in each group (i.e., distributions of the residuals are normal), 2) the population variances in each group are equal (i.e., homogeneity of variances), and 3) observations are independent. In order to use ANOVA we test each group to ensure the three assumptions above hold. For example, for assumption 1) we use the Shapiro-Wilk test, to assess the normality for each group (Shapiro and Wilk, 1965).

If the dependent variables violate the normality assumption, then we use non-parametric statistical analysis that does not require the three assumption above to hold for each group. The next section discusses two non-parametric tests, the Kruskal-Wallis test and the Wilcoxon post-hoc test, that we use to identify the best-performing treatment groups as none of the groups we use satisfy all three assumptions required to use ANOVA.

Kruskal Wallis test

Kruskal and Wallis (1952) offer a non-parametric method for testing whether samples originate from the same distribution. We can use this test to compare two or more independent samples of equal or different sample sizes. A statistically significant Kruskal–Wallis test indicates that at least one sample statistically dominates one other sample. This test compares median values, and its null hypothesis is that the medians of all groups are equal (i.e., the mean ranks of all groups are equal). The alternative hypothesis is that at least one population median of one group is different from the population median of at least one other group. Thus, rejecting the null hypothesis means that the groups are not from identical populations, or at least one of the groups comes from a different population.

The Kruskal–Wallis test is performed for each KPI at a time, and is not performed for all KPIs simultaneously. To perform the Kruskal–Wallis test, we combine all subjects, across all groups (treatment and control), into a single group and sort all subjects in ascending order by the KPI of interest. Second, we assign ranks to the sorted subjects with subjects with the same KPI given the average rank of all tied subjects. For example, a KPI score list of: 50, 100,100, 100, 500 results in a rank list of 1, 3, 3, 3, 5. Third, we add up the different ranks for each group. Then we calculate the H statistic as follows:

$$H = \left[\frac{12}{n(n+1))} \sum_{c}^{j=1} \frac{T_j^2}{n_j} \right] - 3(n+1) \tag{4.2}$$

where n is the total number of subjects across all groups, c is the number of groups, T_j is sum of ranks for the subjects in j^{th} group, and n_j is the number of subjects in the j^{th} group.

After calculating the H statistics, we find the critical chi-square value with $c-1$ degrees of freedom. The value of the H statistic, relative the critical chi-squared values is used to determine if the null hypothesis is rejected. If the critical chi-square value is less than the

H statistic, then we reject the null hypothesis, which is that the medians are equal (i.e., groups are from a same population). If the chi-square value is not less than the H statistic, there is not enough evidence to suggest that the medians are unequal.

The Kruskal–Wallis test does not identify where which groups come from a different populations. Thus, we need to do post-hoc testing such as a Wilcoxon rank sum test to determine if two specific groups' populations are equal.

Wilcoxon rank sum test

The Wilcoxon rank sum test, also known as the Mann-Whitney U-test, offered by Mann and Whitney (1947), is a non-parametric alternative to the two-sample t-test and is used when datasets are skewed or non-normal, as is the case for our study. This test does not require known distribution and examines the difference in median values for two independent groups. The null hypothesis says groups may be generated by sampling from the same populations, and then the alternative says the treatment are not the same or that the medians of the populations sampled to create each group are different. To perform the Wilcoxon test, the sample sizes do not need to be equal.

To calculate Wilcoxon test statistics, first, we do the same as in the Kruskal-Wallis test, we combine all subjects, across all the two groups, into a single group and sort all subjects in ascending order by the KPI of interest. Next, we rank the sorted KPI values the same as in the Kruskal-Wallis test. Finally, the test statistics, $U_{stat} = \min\{U_1, U_2\}$. Where U_1 and U_2 are defined as follows:

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2} \qquad U_2 = R_2 - \frac{n_2(n_2+1)}{2}. \tag{4.3}$$

In (4.3) n_1 and n_2 are the number of subjects in each group, R_1 is the sum of the KPI rankings for group 1, R_2 is defined analogously.

After calculating the U_{stat} test statistic, we need to determine a critical value to help us accept or reject the null hypothesis ($U_{critical}$).

To find the critical value for the Mann Whitney U test, $U_{critical}$, we index the table found in Mann and Whitney (1947) using the values of n_1 and n_2 along with the desired level of statistical significance. If $U_{stat} \leq U_{critical}$ then we reject the null hypothesis and conclude that the two treatment groups are generated from different populations (i.e., populations of two groups are not equal).

4.4 Case Study

Now, we present a case study in which we apply the proposed DDSF to a YouTube company interested in maximizing its ad revenue. First, we discuss the case study's experimental setup introduced in the first stage of the DDSF (i.e., **design of experiment**), and define the dependent and independent variables. Then, we describe the **data collection** methods and provide descriptive statistics of the pre-test dataset along with the outlier detection procedures. Next, we discuss the **running experiment** procedure, including treatment assignment. Finally, we present our empirical results.

4.4.1 Stage 1- Design of Experiment

In this section we discuss the components used in our case study as listed in the **design of experiment** stage found in Section 4.3. We must first define the **subjects** in the experiment. However, prior to defining the subject we must define the platform used. We use YouTube as the platform for our experiment. The subject are videos hosted by an industry partner that has a YouTube channel. We next discuss how videos for this study are selected. The population that is accessible for this case study consists of all YouTube videos uploaded by our industry partner–over 1400 videos. We exclude around 30% of videos from this case study based on their age, popularity and length of videos to create a homogeneous group. Finally, based on the video's age, length and popularity we select 900 videos out of 1400+ videos that follow the following criteria:

- **Age** – New uploaded videos attract more views at the beginning. After a couple of weeks, the viewership falls will back to nominal levels, so all videos for this experiment are at least 30 days age (uploaded one month previously).
- **Length** – Other than single tracks, there are some mixed music videos with a length of 10 minutes and more. We exclude all videos of 10 minutes and more.
- **Popularity** – Some of the channel's videos have gone viral are viewed millions of times per day and might be outliers in our datasets. These outliers produce a long right tail on the distribution of all video's average views. Consequently, to manage outliers, we removed high traffic videos located at the right tail of the average-views distribution.

Next we must define the **independent variables** which are used to define our treatment groups. The independent variables (treatment groups) in this experiment are the types of ads to run on each video. We consider eight different treatment groups that we select from all combinations of ad formats available on the YouTube platform. As in Table 4.2, YouTube offers five ad formats to companies, but not all combinations of ads are possible. For example, *Display ads* are mandatory for all videos. In addition, based on our industry sponsor business requirements, we avoid using *Sponsored cards* in this case study.

Finally we must define the **dependent variables**. After discussing with our industry partner, we converged on two KPIs of interest: 1) **"views"**: The number of legitimate views of a video, and 2) **"gross revenue"**: The estimated gross revenue in USD. Other YouTube companies may want to consider different KPIs such as: daily viewership, watch times, user engagement (e.g., clicks on ads), audience retention, financial performance, etc.

To sum up this section, the experimental setup based on the DDSF is as follows: 1) subjects are YouTube music videos 2) ad formats are the independent variables, and 3) the two dependent variables are views and gross revenue.

4.4.2 Stage 2- Data Collection

We now move to the second stage of the DDSF, Data Collection. In total there are over 1400 videos associated with the channel at the time of this study. In aggregate there are 2.4 billion views times and 7.5 million channel subscribers, as of Ocotober 2019.

The data is collected via the YouTube Analytics and Reporting APIs. Due to the sensitivity of the data collected, we are unable to make the datasets publicly available. To define the data we collect, we must define the data **dimensions**. The time period we use is daily, meaning all data is collected on a daily basis. We also must collect the dependent and independent variable values for each selected time prior. As each treatment group has the independent variable set and unchanged for the duration of the experiment, we collect the dependent variable values (gross revenue and views). In addition to the dependent variables, we collect some control data, such as likes, dislikes, etc. but we find this control data to not impact our results.

Pre-test Datasets Description

Table 4.1 summarizes the statistics of a pre-test dataset. The dataset contains daily *views* and *grossRevenu* of 900 videos for 120 days before running the experiment. The number of observations, N, is $425,700$.

As mentioned above, before running the experiment, we initially excluded some videos based on their age, popularity, and length, to address possible outliers and create a homogeneous group. Additionally, other sources of outliers might appear after we had run the

Table 4.1: Pretest dataset summary statistics - 120 days of data

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
views	425,700	370.426	719.823	0	84	160	363	14,728
gross revenue	425,700	0.689	1.468	0	0.1	0.25	0.65	42.35

experiment. For example, if a blog post refers to one of the experiment's videos, many users might come to the YouTube channel through that blog's link and cause a spike in its daily viewership for a short period. To detect and remove these types of unusual observations, in addition to the nominal data collection mentioned above, we also collect traffic source data via YouTube's APIs. We use the traffic-sources data to detect videos that have a jump in their viewership from an unusual source. In this dataset, we collect the number of daily views grouped by different traffic sources (see Appendix B.1 for the list of traffic sources and their definitions). As a sample, Table B.3 (see Appendix B.3) shows the traffic-source breakdown of all viewers on a specific day for a specific video. Please see Section 4.4.3 that discusses post-test dataset's outlier detection methods in greater detail.

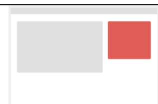
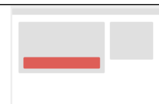
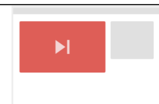
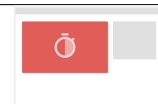
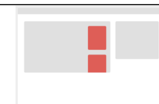
4.4.3 Stage 3- Running Experiment

In this section, we discuss the third stage of our DDSF, **running experiment**, described in general terms in Section 4.3. In the remainder of the section we discuss the components of this stage of the DDSF as applied to our case study in greater detail.

Creating treatment groups

YouTube provides different ad formats that companies can select to show whenever a viewer watches a videos posted by the company. In Table 4.2, the first row contains ad format titles and images that show ad locations and type using red shapes. The second row, Placement, indicates ad locations on a video or watch-page (describes the red shapes in the first row). The next row states each advertisement's format (e.g., text, image, or video). Finally, the last row shows devices that display each ad. We now discuss each ad-format individually in greater detail[3].

Table 4.2: YouTube ad formats

	 **Display ads**	 **Overlay ads**	 **Skippable ads**	 **Non-skippable ads**	 **Sponsored cards**
Placement	In display	In display	In stream	In stream	In display
Format	image/text	image/text	video	video	image/text
Device	Desktop	Desktop	All Devices	Desktop & Mobile	Desktop & Mobile

Display ads: Display ads, which are viewable only in desktop platforms, appear to the right of main videos above the YouTube list of suggested videos to watch next.

Overlay ads: Overlay ads are semi-transparent advertisements that contain a thumbnail, a small picture, and text. The overlay ad is placed such that it covers the lower 20% of the company video.

Skippable ads: Skippable ads (a.k.a. TrueView in-stream ads) are up to 180 seconds in length. They appear at the beginning of YouTube videos to let viewers select if they want to watch the entire ad via a "Skip Ad" button at the screen's right-hand corner. The "Skip Ad" button appears after first five seconds of the ad are played.

Skippable ads are viewable in stream, meaning a viewer must select the play button of the company video before seeing any instream video ads. Skippable ads generate revenue for the company only if a viewer watches the entire ad.

Skippable ads help an advertiser connect with viewers while they engage with videos across YouTube; for example, a 10-minute toy-review video may stop two times to

[3] YouTube ad format table reproduced from https://bit.ly/2QwgrGq accessed on Nov 2019.

display a 30-second video ad, and resume from the same spot after the ad finishes or is skipped by a viewer. Advertisers are interested in viewers watching posted skippable ads as viewers that watch an ad in its entirety are ten times more likely to engage with the brand on YouTube, than other YouTube viewers (Blumenstein, 2015).

A Skippable video ad can be set to play before, during, or after the original video that viewers originally come to view (pre, mid or post roll). A YouTube company may select what type of skippable ads to show, pre/mid/post roll. By default, skippable ads are only set to play pre roll, but a YouTube company may change this setting. In our case study, we do not consider skippable ad type, but it is something to consider in future studies.

Non-skippable ads: Non-skippable ads (a.k.a Standard In-stream ads) are like skippable ads, except the "Skip Ad" button never appears. As a result, non-skippable ads force viewers to watch the entire video ad. Two versions of non-skippable ads exists: short and bumper. Short non-skippable ads, 20 seconds in length, appear before, during or after the main videos. Bumper ads are even shorter video ads, up to 6 seconds, and only appear before the main video.

Sponsored cards: Sponsored cards display ads that may be relevant to the company video and are shown directly over the right-hand side of the company video itself. Sponsored cards appears in-display and show a short "teaser". After a few seconds, the teaser is replaced by a card icon that provides users the choice of clicking through to access more information.

In this research, we use different combinations of the five ad formats listed above as treatments to selected groups. Table 4.3 shows all the treatment groups we consider in this study. For a collection of reasons, outlined below, we do not consider all $2^5 = 32$ ad format combinations. The first row in Table 4.3 shows the treatment groups' names (T_1 to T_8), which are the eight YouTube ad format combinations in this case study. Checkmarks (✓)

in each row mean that we are using that ad format on the selected treatment group. In contrast, crosses (✗) mean that we are not using that ad format on the selected treatment group.

Table 4.3: Treatment groups

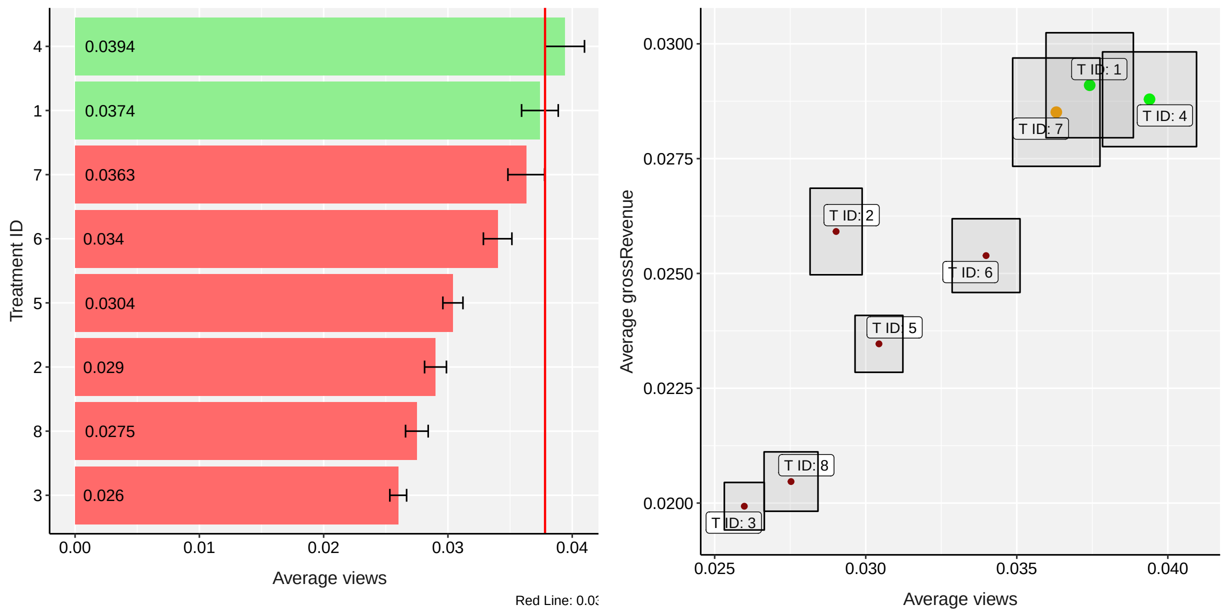
	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8
Display ads	✓	✓	✓	✓	✓	✓	✓	✓
Overlay ads	✓	✓	✓	✓	✗	✗	✗	✗
Skippable video ads	✓	✓	✗	✗	✓	✓	✗	✗
Non-skippable video ads	✓	✗	✓	✗	✓	✗	✓	✗
Sponsored cards	✗	✗	✗	✗	✗	✗	✗	✗

Table 4.3 show that *Display ads* are always check-marked since it is a mandatory ad format for all YouTube videos. Also, do not use *Sponsored cards* in this case study as requested by our industry sponsor, based on their business requirements. We now discuss how subjects, videos, are assigned to each treatment group. We first discuss random assignments, then present the method we use in our case study.

Random treatment assignment

To see how the random assignment works in this case study, we randomly assigned 900 videos to eight different treatment groups (i.e., T_1 to T_8), and draw Figure 4.2 to compare pre-test average *views* and *gross revenues* in the eight treatment groups.

The left panel (4.2a) in Figure 4.2 shows the average *views* per group by eight horizontal bars. Treatment groups that are statistically similar to the treatment group with the highest average views are depicted in green. The y-axis shows treatment IDs (1 to 8,



(a) *views* (b) *views* and *grossRevenu*

Figure 4.2: Random assignment comparison chart- four months of data

listed in descending order of average views), and the x-axis shows the average *views* for each treatment ID on the pre-test dataset (i.e., four-months of observations before the experiment's start date). The error lines, on each bar, are a measure of the precision of the groups' means and show two standard error deviations from the group mean, resulting in a 95% confidence interval. The vertical red line, drawn at the left-hand side of the group with highest average views, is used to visually determine if a treatment group is statistically similar to group with highest average views. Pictorially, so long as the right hand side of the confidence interval of a group is larger than (to the right of) the vertical red line, then those groups are statistically no different than the group with the highest average views at the 95% confidence level. Finally, we can see that random assignment is not perfect and generates non-homogeneous treatment groups since only treatments 4 and 1 are statistically similar on their average *views*.

As we consider two KPIs, we have to construct left panel of Figure 4.2 for *gross revenue*.

However, comparing two graphs may be difficult and instead we combine both graphs into a single graph shown on the righ panel of Figure 4.2. The right panel (4.2b) in Figure 4.2 shows the average *views* on the x-axis and *gross revenue* on the y-axis. This chart contains eight points that indicate a group's average *views* and average *gross revenue* values. The rectangles about each point show a 95% confidence interval for the averages represented by the *views* and *gross revenue* coordinates of the point. The green point shows that Treatment 3 is on a Pareto efficient frontier, which contains those treatments that dominate other, red-point, treatments. Treatment 7, is depicted by a yellow dot, and is a treatment that is not statistically different from Treatment 3. In general, any non-overlapping rectangles between any two treatment groups means that those groups' average *views* and average *gross revenue* are statistically different. The next section introduces the implementation of a GA to this case study to address the shortcoming of random assignments, and generate homogeneous treatment groups.

Genetic Algorithm treatment assignment

We use search algorithms to create homogeneous treatment groups, but finding an exact solution is not always computationally feasible. Thus, we use a Genetic Algorithm (GA), an adaptive heuristic search algorithm, to find a high-quality solution that results in homogeneous treatment groups (i.e., groups that are statistically similar in their average *views* and *gross revenue*.)

Algorithm 1 shows the GA we use for treatment assignment in this case study. The GA, first, randomly generates 100 initial assignments ($nP = 100$). Then, the GA set the probability of mutation to 20 percent ($rM = 0.2$) and the probability of crossover to 80 percent ($rC = 0.8$) to maintain genetic diversity. We run the GA for a maximum of 2000 iterations ($nI = 2000$), and stop the GA when the number of consecutive generations without any improvement reach 500 ($nJ = 500$).

The proposed fitness function (see Algorithm 1) helps us determine how close a specific treatment assignment is to an optimal solution. In other words, all rectangles in the right hand panel of Figure 4.2 need to overlap, meaning treatment groups are homogeneous.

For ease of notation we will refer to the confidence interval rectangles for the two KPIs as *significance rectangles*. An optimal condition for homogeneous groups is all significance rectangles will perfectly overlap. Equivalently, the sum of the distance between the corners (referred to as "cornersDistancesSum") of all significance rectangles will be zero. With this idea in mind we will now present the main procedure of the fitness function ("calculate-Fitness") as follows:

- First, the fitness function receives two parameters: 1) "dateRanges" used to prevent over fitting the data. "dateRanges" is a collection of dates, that span the pre-test dataset, on which we will compute the "cornersDistancesSum" for each range. In this case study we use monthly, bi-monthly, and quarterly date ranges, 2) "length" parameter which is an empty vector that has the "cornersDistancesSum" for each range.

- Second, for each date range, the function subset the pre-test dataset (based on the start-date and the end-date of the selected date range) and calculate average *views* and *gross revenue* for each treatment groups.

- Third, using the standard deviation of *views* and *gross revenue*, the function finds the coordinates of the four corners (i.e., "cornersCoordinates") of the significance rectangle for each treatment group for the selected date range.

- Fourth, the function calculates the total distance between all corners of the significance rectangles for all pairs of treatment groups. The distance between the corners of the significance rectangles for a pair of treatment groups is the "cornersDistances". We now sum the "cornersDistances" for all pairs (non-ordered) of groups, resulting in the “cornersDistancesSum”. The “cornersDistancesSum” is computed for each “dateRange” in “dateRanges.” We aggregate all “cornersDistancesSum” values in a single vector “length” for all “dateRange” values in “dateRanges.”

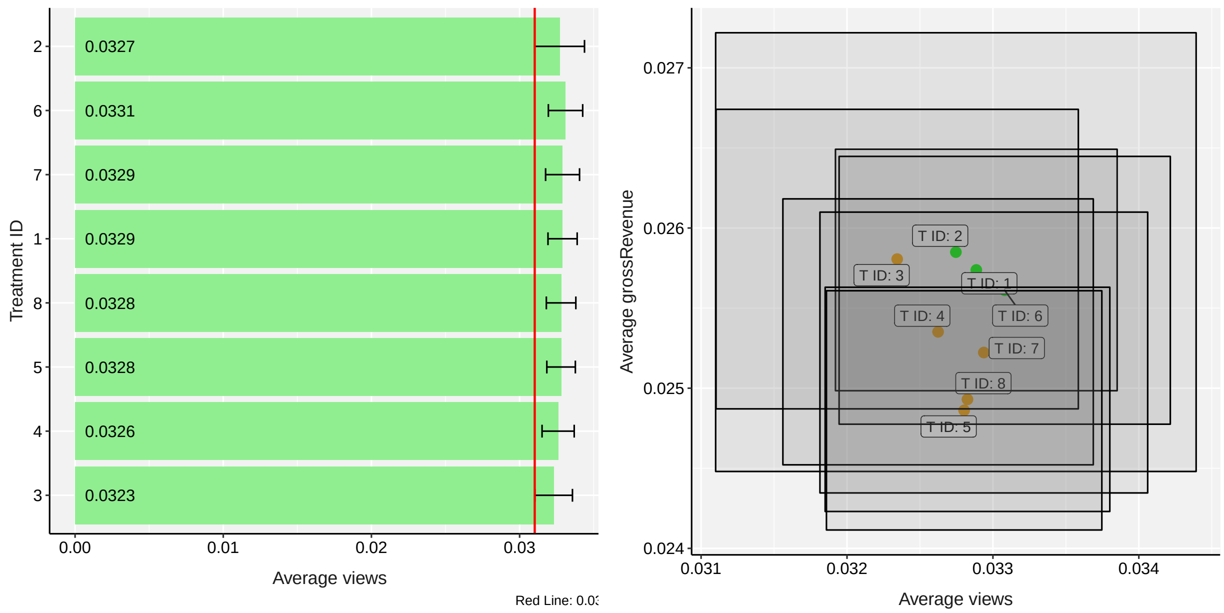
Finally, the function returns the multiplicative inverse of the sum of “length” vector values as the fitness score.

Algorithm 1: Genetic Algorithm (GA) for treatment assignment

Data: *dataSet* - daily views and grossRevenue;
input : nP (population size), rC (crossover rate), rM (mutation rate), nI (iterations), nJ (iterations without fitness score improvement),
output: Solution $X(most\_fit\_member\_of\_population)$

```
initialize(P_I): generate a random population of assignments with size nP
I&J ⟵ 0
while termination_condition_not_met (I ≤ nI or J ≤ nJ ) do
    select assignments from P_I and reproduce rC × nP new assignments
    select two assignments (i.e., parent solutions) from the current population
    crossover parents and form a new assignment (i.e., child)
    if random(0,1) ≤ rM then mutate assignment
    calculateFitness(dataSet, assignment)
    add assignment to the population
    remove the rC × nP least-fit solutions from the population
    I ⟵ I + 1
    if fitness score not improved then J ⟵ J + 1
end
return (most_fit_member_of_population)

Function calculateFitness(dataSet, assignment):
    parameter: dateRanges - a vector of dateRanges
               length - an empty vector of size dateRanges × 1
    foreach dateRange ∈ dateRanges do
        subDataSet ← subsetData(dataSet, dateRange)
        cornersCoordinates ←
         calculateCornersCoord(subDataSet, assignment)
        cornersDistances ← calculateCornersDistances(cornersCoordinates)
        cornersDistancesSum ← sum(cornersDistances)
        lenght[i] ← cornersDistancesSum
    end
    fitness ← 1 / sum(length)
    return fitness
End Function
```



(a) *views* (b) *views* and *grossRevenu*

Figure 4.3: Genetic Algorithm assignment comparison chart- four months of data

Figure 4.3 presents the result of a GA treatment assignment. The left panel shows that all eight treatment groups are in green, meaning that all groups are statistically homogeneous. Also, the right-hand panel shows that all mid-points are close enough that all treatments' significance rectangles overlap. We thus conclude that this GA assignment is appropriate for our case study's experimental design and can use it to introduce treatments and start the experiment.

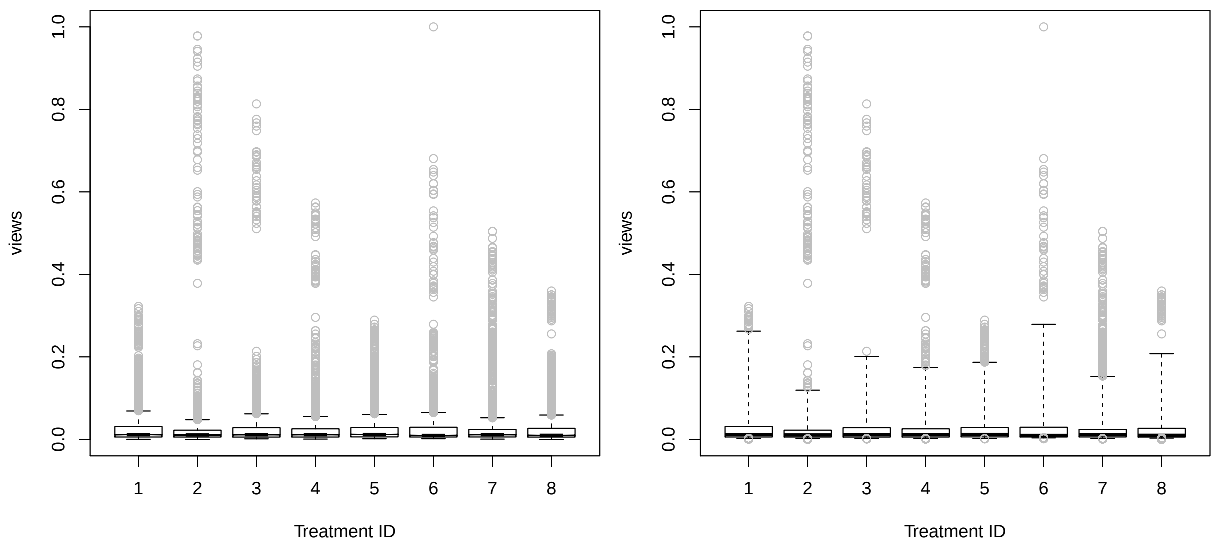
Data processing

After starting the experiment (i.e., assigning case study's subjects to the eight different treatment groups using the GA), we collect the **post-test dataset**. **Outlier detection** procedure is the main component of **data processing** on the third DDSF stage. Outliers are extreme observations that could affect statistical tests comparing treatment groups.

In this case study, we use two methods for detecting outliers. The first method uses the actual number of *views* or *gross revenue* to find possible outliers in each treatment group, using the Adjusted Boxplot tool offered by Hubert and Vandervieren (2008). The second method assesses any unusual behavior or changes in each video's traffic sources. We now discuss each procedure in greater detail.

Adjusted Boxplot

In normally distributed data, the standard Boxplot identifies outliers using interquartile distance ($IQD = Q_3 - Q_1$), which is the difference between the upper quartile (Q_3) and the lower quartile (Q_1). The outlier cutoff interval is $[Q_1 - 1.5 \times IQD\ ;\ Q_3 - 1.5 \times IQD]$. The standard Boxplot, as a popular graphical tool, is not appropriate for detecting outliers for skewed distribution. Hubert and Vandervieren (2008) offer a correction to the outlier cutoff boundaries (i.e., whiskers) to correct the effect of skewness on the standard Boxplot. They use a robust measure of skewness (i.e., Medcouple) in the determination of the cutoff boundaries. Medcouple (MC) is a measure that varies between -1 and 1 and shows a distribution's skewness. For example, a positive value for the MC shows that the distribution is skewed to the right. The adjusted Boxplot includes two functions of MC: $h_l(MC)$ and $h_u(MC)$, in its cutoff values definition. Finally, the adjusted Boxplot cutoff interval is $[Q_1 - h_l(MC) \times IQD\ ;\ Q_3 - h_u(MC) \times IQD]$.



(a) Standard Boxplot

(b) Adjusted Boxplot

Figure 4.4: Outlier detection comparison

Figure 4.4 compares outliers detection of post-test dataset with the standard and adjusted Boxplot charts, in which the x axis shows the number of views and the y axis shows treatment group IDs. In the left chart (4.4a), which is the standard Boxplot, many points exceed the treatment's whiskers and are erroneously declared to be outliers. On the other side (4.4b), which is the adjusted Boxplot graph, the whiskers' lengths are adjusted, and the number of detected outliers is corrected. Because our case study's post-test datasets are skewed (i.e., non-normal), we use an adjusted version of the Boxplot to detect possible outliers. The next section describe the second method to detect and remove outliers.

Traffic-source analysis

To find other possible outliers, we define a two-step procedure for analyzing post-test video-traffic sources. First, we compile a list of suspect-observations by comparing all videos' viewership, grouped by traffic sources, per day. Then, we scrutinize traffic sources of the suspected observations to find any unusual changes in each video's viewership traffic source.

In the first step, for each video we determine the number of views that come from each traffic source. In this case study we have 16 traffic sources, thus for each day we generate a 16 dimensional vector with the number of views coming from each source. For each day, t, we determine the traffic source vector, $\mathbf{TS}_t$. To identify large changes in traffic source for day t we compute the Pearson Correlation Coefficient (PCC) between $\mathbf{TS}_t$ and $\mathbf{TS}_{t-1}$. The computed PCC tells us the linear correlation between the two adjacent days' traffic sources. If the PCC is near one, then it means that there are no significant changes in the video-traffic sources. However, if the PCC is less than 0.98, then we flag that observation to analyze the source of variation caused by any changes in the traffic sources.

For each flagged observation, we calculate the PCC for the same video over last 14 days (pairwise dates, i.e., we compute the PCC for the following set of vectors: $\{(\mathbf{TS}_t, \mathbf{TS}_{t-1}),$ $(\mathbf{TS}_{t-1}, \mathbf{TS}_{t-2}), \ldots, (\mathbf{TS}_{t-13}, \mathbf{TS}_{t-14})\}$). Then we calculate the mean and standard deviation (SD) of the PCCs history to see what is the distance (in terms of SD) of the flagged PCC with its PCCs historic two-week mean. Now, we select those observations that have three or more SD distances from the PCCs mean. All these selected observations (i.e., observations for a specific video on a specific day) are suspect to be an outlier. A shift in traffic sources does not mean that there was external action taken, e.g., a blog post. Instead, it might be due to endogenous YouTube action, such as posting the video near the top of a viewer's recommended videos to watch next. As such, we analyze the traffic sources of the suspected videos to assess whether the traffic sources are external to YouTube (e.g., blog post links, other video referrals, and social media links).

The following list shows external traffic sources that have been identified by YouTube[4]:

- **NO_LINK_OTHER**: This category encompasses direct traffic to a video. However,

[4] External traffic sources' definitions are reproduced from YouTube API documentations: https://bit.ly/37njGFN accessed on Nov 2019.

YouTube did not identify a referrer for the traffic.

- **EXT_URL**: The video views were referred from a link on another website.
- **NO_LINK_EMBEDDED**: This category encompasses direct traffic to a video as well as traffic on mobile apps.
- **ANNOTATION**: Viewers reached the video by clicking on an annotation in another video.
- **CAMPAIGN_CARD**: Views originated from claimed, user-uploaded videos that the content-owner used to promote the viewed content.

Continuing the outlier detection procedure, we check whether the number of views through each of the five traffic sources listed above increased by 50 views. The value of 50 allows us to account for videos that have a small number of daily views. For example, if a video has 10 views per day a natural variation of 2 additional views from a particular source results in a total of 50% of all views coming from that source we do not flag this video as an outlier. However, 50 views may be too small for videos with a larger number of views per day. To address this issue we consider the total percent of views coming from each source. If a source is less than 5% of views for a given video, then any change in number of views for that source is ignored, regardless of its value. For example, if the EXT_URL traffic source of a suspected observation increased by 250 views on that day and the percentage of EXT_URL traffic source is 4% of all views, then we conclude that the video is not an outlier. Finally, the traffic source analysis on the experiment datasets shows that the percentage of possible outliers detected by this method is less than two percent of the experiment's total observations.

4.4.4 Stage 4- Making Decision

The fourth stage of the DDSF, **making decision**, provides insights to our industry-partner on the treatments that perform best. This case study uses a multi-rounds strategy for running the experiment. A multi-round approach is used to minimize the cost of finding the best treatment group.

Each round of the experiment presents non-parametric statistical tests results based on two dataset versions: 1) dataset including outliers, and 2) dataset without outliers. It is important to exclude outliers to make sure statistical tests are correct; however, and at the same time, considering outliers is useful for business decisions since outliers are a part of the firm's asset portfolio. The next section discusses the process of **comparing treatment groups**, using a multi-round experiment strategy, to **find best groups** that generate the largest revenue.

Comparing treatment groups

Using the multi-round experiment strategy, first, we include all possible treatment groups in the first round of the experiment (i.e., T_1 to T_8). Analyzing the result of the first round identifies treatment groups that are performing better in terms of generating more views and gross revenue. In the second round, we only compare the best performing treatment groups and start a new round of experiment with those high-performing groups. Please note that selecting high-performing groups may be done statistically, but we suggest to use both quantitative and decision maker intervention.

After each round of the experiment we run the Kruskal-Wallis test on the two post-test datasets, one with and one without outliers. In addition, using the two post-test datasets we create two Pareto efficiency charts to help determine the best treatment groups. Pareto efficiency rules are helpful to find dominant treatment groups when more than one KPI is considered in the experimental study. However, in the case of considering one KPI, the Wilcoxon post-hoc rank-sum test may be used to determine the best treatment group. In this case study, we do not use the Wilcoxon test results in the decision-making process because we use two KPIs in our experimental setup (see Appendix B.2 for more information on one-tailed Wilcoxon test results). The remainder of this section introduces the results of two rounds of the experimental case study.

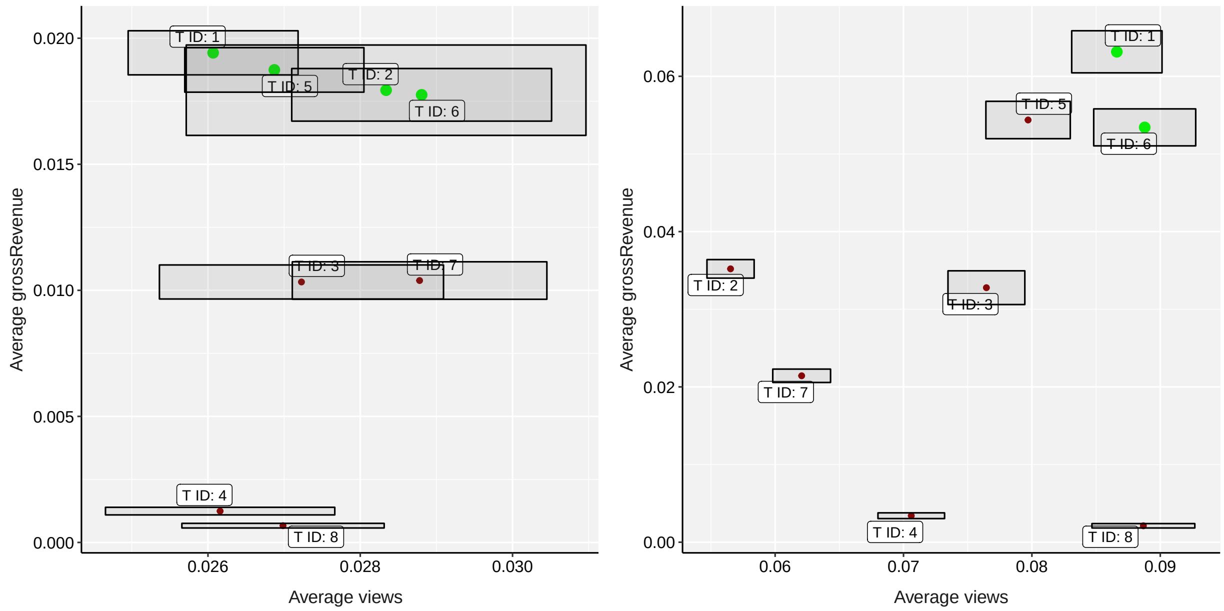
First round

We start the first round of the experiment by introducing the eight treatment groups (see Table 4.3) to 896 videos, and then collected daily video data of KPIs of interest over 40 days. As discussed in Section 4.3.2, the Kruskal-Wallis non-parametric analysis of variance helps to discover treatment groups that are significantly different from the others. Table 4.4 shows the Kruskal-Wallis test statistics for the first round experiment. For example, the first row (views) shows that the chi-squared (χ^2) test statistics, H, is 410.91 for the original dataset including outliers, and the test statistic for the dataset without possible outliers is 360.12. In this table, the degrees of freedom and p-values are 7 and $< 2.2\text{e}{-16}$, respectively. Thus, when the p-values are less than five percent, we conclude that the different ad formats on YouTube videos significantly affected views and gross revenue, and the distributions of treatment groups for both KPIs are statistically different (i.e., treatment groups are not from identical populations).

Table 4.4: Kruskal-Wallis test - First round

KPIs	chi-squared (dataset with outliers)	chi-squared (dataset without outliers)
views	410.91 ($df = 7,$ p-value $\leq 2.2e - 16$)	360.12 ($df = 7,$ p-value $\leq 2.2e - 16$)
gross revenue	2379.1 ($df = 7,$ p-value $\leq 2.2e - 16$)	2136.9 ($df = 7,$ p-value $\leq 2.2e - 16$)

Figure 4.5 shows the average views and gross revenue for all videos grouped by the eight treatments at the same time. The left panel (4.5a) is based on original dataset (i.e., including outliers), and the right panel (4.5b) is based on a dataset without outliers.



(a) Dataset with outliers

(b) Dataset without outliers

Figure 4.5: The first round experiment's results

The left panel of Figure 4.5 shows that treatment groups 1, 2, 5, and 6 are high-performing groups, and dominate the other groups (i.e., the low-performing groups) considering views and gross revenue performance. Also, in the right panel, from which we removed possible outliers, treatment groups 1, 6 and 5 dominate the other treatment groups. We then conclude that treatment groups 1, 2, 5, and 6 are performing better in this round, so we stop the first round of the experiment by setting the ad format to the default group (i.e., treatment 1), and run the next round of the experiment only with the high-performance treatment groups.

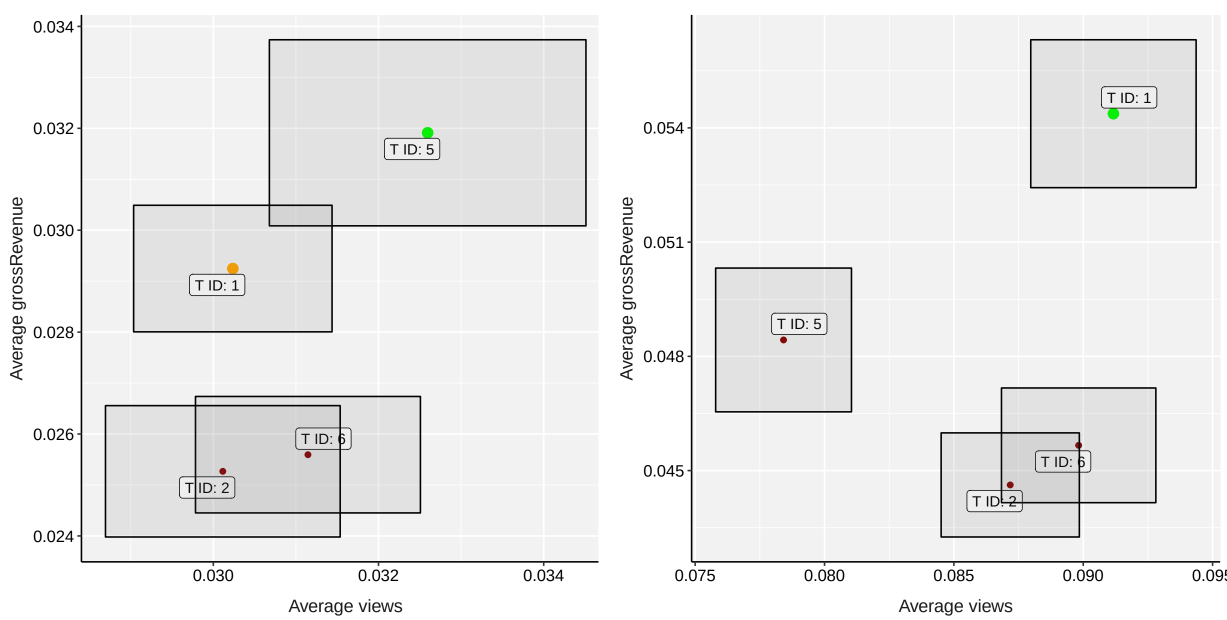
Second round

The second round of the experiment includes high-performance treatment groups from the first round: treatments 1, 2, 5 and 6. To start the second round, we waited for 60 days after stopping the first round by setting back on all videos' ad formats to the default configuration (i.e., Treatment 1). Then, we ran the GA again to assign all videos to the top-performing groups, and started to collect a new post-test dataset for the second round of the experiment.

Table 4.5: Kruskal-Wallis test - Second round experiment

KPIs	chi-squared (dataset with outliers)	chi-squared (dataset without outliers)
views	284.14 $(df = 3,\ \text{p-value} \leq 2.2e - 16)$	257.56 $(df = 3,\ \text{p-value} \leq 2.2e - 16)$
gross revenue	122.45 $(df = 3,\ \text{p-value} \leq 2.2e - 16)$	109.46 $(df = 3,\ \text{p-value} \leq 2.2e - 16)$

Table 4.5 shows the Kruskal-Wallis test statistics for the second round, in which all p-values are near zero. Thus, we conclude that treatment groups in the second round of the experiment differ from each other.



(a) Dataset with outliers

(b) Dataset without outliers

Figure 4.6: The second round experiment's results

The left panel of Figure 4.6 shows that treatment groups T_5 and T_1 are Pareto efficient using the original post-test dataset (dataset with outliers) in the second round of the experiment. T_5 (the best group in the left panel) is similar to T_1, except it does not include the Overlay ads. Thus, it would seem to suggest that the exclusion of Overlay ads from a treatment group may increase views and gross revenue simultaneously. Also, this fact is true when we compare T_2 and T_6. Since, T_6 ad formats is similar to T_2, except that T_6 has not Overlay ads. Moreover, the comparison of pairs $[T_2, T_1]$ and $[T_6, T_5]$ shows that including non-skippable ads has no significant effect on views, but could increase gross revenues.

The right panel (4.6b) shows that the treatment T_1 is the best treatment group (dataset without outliers). Thus, we conclude that the treatment group T_1, which includes all four admissible ad formats, is the best group in terms of maximizing gross revenue and viewership. However, given the natural evolution of YouTube viewership, it is likely best to continuously test candidate ad formats to see if viewers' response to ad formats changes over time.

Finally, to make sure our large sample sizes do not affect our analysis based on the tables 4.5 and 4.4, we randomly sub-sampled both datasets (i.e., with and without outliers) 100 times. The results of the Kruskal Wallis tests with the sub-sampled datasets (see Table B.5 and B.4 in Appendix B.4) show that the significance of p-values with the smaller samples are the same for both rounds (i.e., sub-sampled KW tests' p-values are less than five percent).

4.5 Conclusions

To investigate how companies can boost their revenue from online advertisement platforms, we developed a new data-driven statistical framework (DDSF) to determine which type of ads to show in order to maximize YouTube revenue. Different combinations of ad formats generate different revenues. One way to find the ad format that maximizes revenue is to simply test all ad formats, and select the one that generates the largest revenue. However, this experiment may lead to substantial losses for the YouTube company.

Our proposed DDSF is based on a quasi-experimental design that uses a genetic algorithm for treatment assignment and finds the best ad format combination. The statistical framework performs non-parametric analysis, such as the Kruskal-Wallis and Wilcoxon post-hot test. Finally, it provides a near-real-time statistical process to choose the best set of ad formats YouTube companies may use on their videos.

We apply the DDSF to a YouTube company to show which combination of ad formats maximize its YouTube revenue. YouTube provides different ad formats for YouTube companies to choose from, each with its own properties, see Table 4.2.

The result of the case study is the revenue-maximizing ad format combination (i.e., the list of ad format types a firm should use to maximize its ad revenue) for our industry partner. The case study's findings are summarized in the following main points:

- Random assignment does not create homogeneous treatment groups. The proposed GA is one way to generate homogeneous treatment groups.
- In the first round of the experiment, four treatment groups are identified as high-performance groups (i.e., T_1, T_2, T_5 and T_6).
- Skippable ads are not present in any of the low-performing groups (i.e. T_3, T_4, T_7 and T_8). This fact suggests that the Skippable ads format has a significant effect on gross revenue, and not a significant negative impact on average views.
- In the second round of the experiment, two treatments (i.e., T_1 and T_5) are identified as the best groups based on original post-test dataset (the dataset with outliers). By

removing the possible outliers, the case study's experiment results in treatment T_1 as the best treatment group that include all admissible ad formats (i.e., Display ads, Overlay ads, Skippable video ads, and Non-skippable video ads).

- The second round of the experiment suggests that not choosing the Overlay ads could increase views and gross revenue simultaneously, at least when not considering outliers. It also implies that the Non-skippable ads has no significant effect on views, but may increase gross revenues.

- The difference between our results for the post-test datasets with and without outliers suggests that there is "inelastic" demand for certain YouTube videos. What we mean by "inelastic" is independent of the type of ad shown to a viewer, if the viewer really wants to see the video, then the viewer will view the ad. We are led to this conclusion by seeing the change in performance of treatment group 5 in the second round of the experiment when comparing the datasets with and without outliers.

The proposed DDSF is a needed first step in helping YouTube companies manage their YouTube channels. However, the DDSF is not without its shortcoming. For example, the DDSF does not scale well if there are many independent variables, relative to the number of subjects. In addition, a YouTube firm may need an optimization based approach to finding the profit maximizing policy, something that the framework does not explicitly consider. Finally, the DDSF was run over a fixed time period. The generalizability of this study is limited, because it is well known that YouTube viewership changes not only diurnally, but also annually, and a robustness to the natural annual fluctuations of YouTube viewership needs to be considered when finding the best ad formats to show on a video. Theoretically, it may be that one set of ad formats are best to show in the summer and another set in the winter. However, the proposed DDSF offers a general statistical framework that can be run at a different time and also applies to all kinds of online platforms for running an advertisement experiment.

The above shortcoming open up a large set of potential future studies. The first future research direction is to apply the current DDSF and experimental setup to multiple YouTube companies. Currently, it is unclear if the endogenous attributes of the YouTube company we worked with lead to the observed results. Second, using the same DDSF, but consider additional independent variables such as in-stream ad location. Finally, an optimization based experimental framework may be used to mitigate costs and find a near-optimal set of ad formats to show on YouTube videos. This approach may lead to loss of statistical accuracy, but at the gain of lower opportunity costs.

Chapter 5

Conclusion

Business Analytics (BA) helps organizations improve their business by making optimized decisions with the use of statistical tools that require data collection and processing. For example, marketing communities often use Business Analytics in predicting and analyzing consumer behavior. However, without high-quality data and statistics, Business Analytics have little meaning to any organization. Thus, high-quality data and statistical tools are the two primary components of any BA system.

In this dissertation, all three essays use both primary components of BA (i.e., real-world business data and statistical tools) to address three business problems in the context of financial market and online advertisement industry. The first essay uses two different datasets: 1) the US prediction market data and 27 different global financial markets index data. It also uses time-series analysis to find causal relationships between the US prediction markets and the financial market indexes. The other two essays used various online advertisement datasets (collected from YouTube music channels) and used panel data analysis, experimental design and non-parametric statistical analysis to addressed two business problems for an online advertisement company.

Using a data-driven statistical approach, Business Analytics has a wide range of applications that are categorized into three types: descriptive, predictive, and prescriptive analytics. Descriptive analytics use computer science and statistical techniques to understand the effectiveness of previous business decisions (i.e., past performance) and provide insights to answer "what happened?" Predictive Analytics uses statistical models and forecasting techniques to understand the future and answer "what will happen?" Finally, prescriptive analytics utilize optimization techniques to recommend actions a business may take to improve performance and answer, "what should we do?". The essays presented in this dissertation are in the predictive and prescriptive analytics categories as follows:

- The first essay provides insight into how prediction market data may be used as a predictive tool to understand reasons for global financial markets' fluctuations. Running real-time causality analysis may help traders understand/analyze financial market behaviours resulting from political events, such as U.S. presidential elections.
- The second essay, which is in the realm of predictive analytics, helps YouTube content owners understand what may happen with their viewership when they choose a specific IP protection policy. This essay clarifies for our YouTube partner what is the impact of UGC video viewership on the YouTube partner's channel viewership under two different YouTube IP protection policies. The result of this study can be generalized to any YouTube music channel, which is one of the top four content categories watched by YouTube users (Google, 2016*d*). Our results are a needed first step in helping a YouTube channel determine what IP protection policy is best suited for its videos.
- The third essay, which is a prescriptive analytics study, provides a data-driven statistical framework (DDSF) to determine which type of ads to show in order to maximize YouTube revenue. Though the decision of what type of ads to show on a video is not automated in our study, the tools exists for our YouTube partner to automatically set and update advertisement types automatically over time. Based on the DDSF we designed an experiment for a YouTube company to show which combination of

ad formats maximize their revenue. The result of the experiment shows the revenue-maximizing ad format combination (i.e., the list of ad format types a firm should use to maximize its ad revenue). Despite the generalizability limitations of this study, the proposed DDSF offers a general statistical framework that applies to all kinds of online platforms for running an advertisement experiment.

BA is a rapidly growing field taking a problem-first approach in developing and understanding solutions. In the future, we will continue applying the methods learning in this dissertation to various other real-world problem. In addition, to these methods we will further integrate methods from Artificial Intelligence (AI), Machine Learning (ML), and Operations Research (OR). For example, some current work is to use prediction market data to find optimal trading strategies in such markets or financial markets.

References

Abe, N. (1999), Learning to optimally schedule internet banner advertisements, *in* 'Proc. of 16th Int. Conf. on Machine Learning', pp. 12–21.

Adobe (2015), 'Seo for success in video marketing — adobe', https://goo.gl/W2GDo9. (Accessed on 09/16/2017).

Agarwal, D., Chen, B.-C. and Elango, P. (2009), Explore/exploit schemes for web content optimization, *in* '2009 Ninth IEEE International Conference on Data Mining', IEEE, pp. 1–10.

AIB (2018), 'Iab internet advertising revenue report, 2018 full year results'.
URL: *https://bit.ly/2O8rxjf*

Akaike, H. (1973), Information theory and an extension of the maximum likelihood principle, *in* B. N. Petrov and F. Csaki, eds, 'Second International Symposium on Information Theory', Akadémiai Kiado, Budapest, pp. 267–281.

Aktas, H. and Oncu, S. (2006), 'The stock market reaction to extreme events: the evidence from turkey', *International Research Journal of Finance and Economics* **6**(6), 78–85.

Alesina, A. F. and Sachs, J. D. (1986), 'Political parties and the business cycle in the united states, 1948-1984'.

Alesina, A. and Perotti, R. (1996), Budget deficits and budget institutions, Technical report, National Bureau of Economic Research.

Alesina, A. and Rosenthal, H. (1995), *Partisan politics, divided government, and the econ-*

omy, Cambridge University Press.

Alesina, A. and Roubini, N. (1992), 'Political Cycles in OECD Economies', *Source: The Review of Economic Studies* **59**(4), 663–688.

Allen, A. (2007), 'Battling in the name of balance: Evaluating solutions to copyright conflict in viacom international v. youtube', *BYU L. Rev.* p. 1023.

Allvine, F. C. and O'Neill, D. E. (1980), 'Stock market returns and the presidential election cycle: Implications for market efficiency', *Financial Analysts Journal* **36**(5), 49–56.

Araman, V. and Fridgeirsdottir, K. (2011), A uniform allocation mechanism and cost-per-impression pricing for online advertising, Technical report, Working paper.

Armstrong, J. (2003), 'Digital-rights management: Business and technology.'.

Arora, V. and Vamvakidis, A. (2004), 'The impact of us economic growth on the rest of the world: How much does it matter?', *Journal of Economic Integration* pp. 1–18.

Auer, P., Cesa-Bianchi, N. and Fischer, P. (2002), 'Finite-time analysis of the multiarmed bandit problem', *Machine learning* **47**(2-3), 235–256.

Aylward, A. and Glen, J. (2000), 'Some international evidence on stock prices as leading indicators of economic activity', *Applied Financial Economics* **10**(1), 1–14.

Baker, S. R., Bloom, N. and Davis, S. J. (2016), 'Measuring economic policy uncertainty', *The quarterly journal of economics* **131**(4), 1593–1636.

Baltagi, B. H. and Chang, Y.-J. (1994), 'Incomplete panels: A comparative study of alternative estimators for the unbalanced one-way error component regression model', *Journal of Econometrics* **62**(2), 67–89.

Balvers, R. J., Cosimano, T. F. and McDonald, B. (1990), 'Predicting stock returns in an efficient market', *The Journal of Finance* **45**(4), 1109–1128.

Bartels, L. M. (1988), *Presidential primaries and the dynamics of public choice*, Princeton University Press.

Bartholomew, T. B. (2014), 'The death of fair use in cyberspace: Youtube and the problem with content id', *Duke L. & Tech. Rev.* **13**, 66.

Bawn, K. and Rosenbluth, F. (2006), 'Short versus long coalitions: electoral accountability and the size of the public sector', *American Journal of Political Science* **50**(2), 251–265.

BBC (2007*a*), 'Viacom will sue youtube for $1bn', https://goo.gl/VqdWd2. (Accessed on 08/28/2017).

BBC (2007*b*), 'Youtube facing football lawsuit', https://goo.gl/4LQ1t3. (Accessed on 08/28/2017).

Bell, R. and McCaffrey, D. (2002), 'Bias reduction in standard errors for linear regression-with multi-stage samples.', *Statistics Canada* .

Belobaba, P. (1987), Air travel demand and airline seat inventory management, PhD thesis, Massachusetts Institute of Technology.

Belobaba, P. P. (1989), 'Or practice—application of a probabilistic decision model to airline seat inventory control', *Operations Research* **37**(2), 183–197.

Berg, J. E., Nelson, F. D. and Rietz, T. A. (2008), 'Prediction market accuracy in the long run', *International Journal of Forecasting* **24**(2), 285–300.

Berg, J. E. and Rietz, T. A. (2003), 'Prediction markets as decision support systems', *Information systems frontiers* **5**(1), 79–93.

Berg, J., Nelson, F. and Rietz, T. (2003), 'Accuracy and forecast standard error of prediction markets', *Tippie College of Business Administration, University of Iowa* .

Berg, M. (2019), 'The highest-paid youtube stars of 2019: The kids are killing it'.
URL: *https://bit.ly/2wADEzD*

Berk, R. A., Smyth, G. K. and Sherman, L. W. (1988), 'When random assignment fails: Some lessons from the minneapolis spouse abuse experiment', *Journal of Quantitative Criminology* **4**(3), 209–223.

Bernstein, F., Kök, A. G. and Xie, L. (2015), 'Dynamic assortment customization with limited inventories', *Manufacturing & Service Operations Management* **17**(4), 538–553.

Białkowski, J., Gottschalk, K. and Wisniewski, T. P. (2008), 'Stock market volatility around national elections', *Journal of Banking & Finance* **32**(9), 1941–1953.

Binder, M., Hsiao, C. and Pesaran, M. H. (2005), 'Estimation and inference in short panel vector autoregressions with unit roots and cointegration', *Econometric Theory* **21**(4), 795–837.

Bitran, G. and Caldentey, R. (2003), 'An overview of pricing models for revenue management', *Manufacturing & Service Operations Management* **5**(3), 203–229.

Bitran, G. R. and Gilbert, S. M. (1996), 'Managing hotel reservations with uncertain arrivals', *Operations Research* **44**(1), 35–49.

Bitran, G. R. and Mondschein, S. V. (1997), 'Periodic pricing of seasonal products in retailing', *Management science* **43**(1), 64–79.

Bittlingmayer, G. (1998), 'Output, stock volatility, and political uncertainty in a natural experiment: Germany, 1880–1940', *The Journal of Finance* **53**(6), 2243–2257.

Blumenstein, H. (2015), 'Google august 2015 meta-analysis: Measuring trueview impact on brand channel engagement', https://www.thinkwithgoogle.com/marketing-resources/how-youtube-extends-reach-engagement-of-video-advertising/. (Accessed on 09/15/2019).

Borge, A. (2015), '6 factors shaping the global economy in 2016', https://www.weforum.org/agenda/2015/12/6-factors-shaping-the-global-economy-in-2016/. (Accessed on 09/27/2018).

Borges, M. R. (2010), 'Efficient market hypothesis in european stock markets', *The European Journal of Finance* **16**(7), 711–726.

Boroughf, B. (2015), 'The next great youtube: Improving content id to foster creativity, cooperation, and fair compensation', *Alb. LJ Sci. & Tech.* **25**, 95.

Breen, W., Glosten, L. R. and Jagannathan, R. (1989), 'Economic significance of predictable variations in stock index returns', *The Journal of Finance* **44**(5), 1177–1189.

Breusch, T. S. (1978), 'Testing for autocorrelation in dynamic linear models*', *Australian Economic Papers* **17**(31), 334–355.

Breusch, T. S. and Pagan, A. R. (1980), 'The lagrange multiplier test and its applications to model specification in econometrics', *The Review of Economic Studies* **47**(1), 239–253.

Brown, K. C., Harlow, W. and Tinic, S. M. (1988), 'Risk aversion, uncertain information, and market efficiency', *Journal of Financial Economics* **22**, 355–385.
URL: *https://faculty.mccombs.utexas.edu/keith.brown/Research/JFE-12.88.pdf*

Brüggelambert, G. (2004), 'Information and efficiency in political stock markets: using computerized markets to predict election results', *Applied Economics* **36**(7), 753–768.

Buhse, W. (2001), 'Digital rights management for music filesharing communities', *AMCIS 2001 Proceedings* p. 296.

Cahan, J., Malone, C. B., Powell, J. G. and Choti, U. W. (2005), 'Stock market political cycles in a small, two-party democracy', *Applied Economics Letters* **12**(12), 735–740.

Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2011), 'Robust inference with multiway clustering', *Journal of Business & Economic Statistics* **29**(2), 238–249.

Cameron, A. C. and Trivedi, P. K. (2010), *Microeconometrics using stata*, Vol. 2, Stata press College Station, TX.

Campbell, D. T. and Cook, T. D. (1979), *Quasi-experimentation: Design & analysis issues for field settings*, Rand McNally College Publishing Company Chicago.

Carlsen, F. (2000), 'Unemployment, inflation and government popularity—are there partisan effects?', *Electoral Studies* **19**(2-3), 141–150.

Carroll, W. J. and Grimes, R. C. (1995), 'Evolutionary change in product management: Experiences in the car rental industry', *Interfaces* **25**(5), 84–104.

Chen, B., Yuan, S. and Wang, J. (2014), A dynamic pricing model for unifying programmatic guarantee and real-time bidding in display advertising, *in* 'Proceedings of the Eighth International Workshop on Data Mining for Online Advertising', ACM, pp. 1–9.

Chi, C. (2019), '51 youtube stats every video marketer should know in 2019'.
URL: *https://blog.hubspot.com/marketing/youtube-stats*

Chickering, D. M. and Heckerman, D. (2003), 'Targeted advertising on the web with inventory management', *Interfaces* **33**(5), 71–77.

Chiu, C.-L., Chen*, C.-D. and Tang, W.-W. (2005), 'Political elections and foreign investor trading in south korea's financial markets', *Applied Economics Letters* **12**(11), 673–677.

Chorianopoulos, K., Barria, J., Regner, T. and Pitt, J. (2005), Cross media digital rights management for online stores, *in* 'First International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution (AXMEDIS'05)', IEEE, pp. 4–pp.

Ciancimino, A., Inzerillo, G., Lucidi, S. and Palagi, L. (1999), 'A mathematical programming approach for the solution of the railway yield management problem', *Transportation science* **33**(2), 168–181.

Cisco (2017), 'Cisco visual networking index: Forecast and methodology, 2016–2021 - cisco', https://goo.gl/zu9fze. (Accessed on 09/16/2017).

Clement, J. (2020), 'Google: quarterly revenue 2019'.
URL: *https://www.statista.com/statistics/267606/quarterly-revenue-of-google/*

Cloak, T. (2007), 'Digital titanic: The sinking of youtube. com in the dmca's safe harbor, the', *Vand. L. Rev.* **60**, 1559.

Collings, C. (2008), 'Viacom v. youtube: The crossroads of copyright law', *Brigham Young University Prelaw Review* **22**(1), 7.

Collins, S. (2014), 'Youtube and limitations of fair use in remix videos', *Journal of Media Practice* **15**(2), 92–106.

Colombo, V. (2013), 'Economic policy uncertainty in the us: Does it matter for the euro area?', *Economics Letters* **121**(1), 39–42.

Comincioli, B. (1996), 'The stock market as a leading indicator: An application of granger causality', *University avenue undergraduate journal of economics* **1**(1), 1.

Cooray, A. and Wickremasinghe, G. (2007), 'The efficiency of emerging stock markets: Empirical evidence from the south asian region', *The Journal of Developing Areas* pp. 171–183.

Cope, B. and Freeman, R. (2001), *Digital rights management and content development*, Common Ground.

Cornell (2017), '17 u.s. code § 512 - limitations on liability relating to material online', https://goo.gl/yXz69w. (Accessed on 08/28/2017).

Craver, S., Wu, M., Liu, B., Stubblefield, A., Swartzlander, B., Wallach, D. S., Dean, D. and Felten, E. W. (2001), 'Reading between the lines: Lessons from the sdmi challenge.'.

Davidson, R. and MacKinnon, J. G. (2004), *Instructor ' s Manual to Accompany Econometric Theory and Methods*, Oxford University Press.

Davies, D. (2018), 'The 7 most popular search engines in the world - seo 101'.
URL: *https://www.searchenginejournal.com/seo-101/meet-search-engines/close*

Davis, S. J. (2016), An index of global economic policy uncertainty, Technical report, National Bureau of Economic Research.

De Bondt, W. F. and Thaler, R. (1985), 'Does the stock market overreact?', *The Journal of finance* **40**(3), 793–805.

De Haan, J. and Sturm, J.-E. (1994), 'Political and institutional determinants of fiscal policy in the european community', *Public choice* **80**(1-2), 157–172.

DeBondt, W. F. M. and Thaler, R. H. (1985), 'Does the Stock Market Overreact?', *Journal of Finance* **40**(3), 793–805.

Dickey, D. A. and Fuller, W. A. (1979), 'Distribution of the Estimators for Autoregressive Time Series With a Unit Root', *Journal of the American Statistical Association* **74**(366), 427–431.

Diks, C. and Panchenko, V. (2006), 'A new statistic and practical guidelines for nonparametric granger causality testing', *Journal of Economic Dynamics and Control* **30**(9-10), 1647–1669.

Döpke, J. and Pierdzioch, C. (2006), 'Politics and the stock market: Evidence from germany', *European Journal of Political Economy* **22**(4), 925–943.

Driscoll, M. T. (2007), 'Will youtube sail into the dmca's safe harbor or sink for internet piracy?', *Review of intellectual property law* .

Dufour, J.-M. and Renault, E. (1998), 'Short run and long run causality in time series: theory', *Econometrica* pp. 1099–1125.

Economics, F. (2011), 'Estimating the global economic and social impacts of counterfeiting and piracy', *A report commissioned by business action to stop counterfeiting and piracy (BASCAP), An ICC Initiative* .

Eff (2017), 'Viacom v. youtube — electronic frontier foundation', https://goo.gl/Y3ZS4v. (Accessed on 08/26/2017).

Engle, R. F. and Granger, C. W. J. (1987), 'Co-integration and Error Correction: Representation, Estimation, and Testing', *Econometrica* **55**(2), 251–76.

Erikson, R. S. (1989), 'Economic Conditions and The Presidential Vote', *American Political Science Review* **83**(2), 567–573.

Fair, R. (1978), 'The Effects of Economic Events on Votes for Presidents', *Review of Economics and Statistics* **60**(2), 159–172.

Fama, E. (1991), 'Efficient markets: Ii, fiftieth anniversary invited paper', *Journal of Finance* **46**, 1575–1617.

Fama, E. F. and Malkiel, B. G. (1970), 'Efficient capital markets: A review of theory and empirical work', *The journal of Finance* **25**(2), 383–417.

Feng, Y. and Gallego, G. (1995), 'Optimal starting times for end-of-season sales and optimal stopping times for promotional fares', *Management science* **41**(8), 1371–1391.

Feng, Y. and Gallego, G. (2000), 'Perishable asset revenue management with markovian time dependent demand intensities', *Management science* **46**(7), 941–956.

Ferson, W. E. and Harvey, C. R. (1993), 'The risk and predictability of international equity returns', *Review of financial Studies* **6**(3), 527–566.

Fetscherin, M. (2002), 'Present state and emerging scenarios of digital rights management systems', *International Journal on Media Management* **4**(3), 164–171.

Fetscherin, M. and Schmid, M. (2003), Comparing the usage of digital rights management systems in the music, film, and print industry, *in* 'Proceedings of the 5th international conference on Electronic commerce', ACM, pp. 316–325.

Fiorina, M. P. (1991), Elections and the economy in the 1980s: Short-and long-term effects, *in* 'Politics and Economics in the Eighties', University of Chicago Press, pp. 17–40.

Foerster, S. R. (1993), 'Stock market performance and elections: made-in-canada effects?', *Papers in Political Economy* .

Foerster, S. R. and Schmitz, J. J. (1997), 'The transmission of us election cycles to international stock returns', *Journal of International Business Studies* **28**(1), 1–13.

Foroughi, A., Albin, M. and Gillard, S. (2002), 'Digital rights management: a delicate balance between protection and accessibility', *Journal of information science* **28**(5), 389–395.

Fridgeirsdottir, K. and Najafi-Asadolahi, S. (2009), Revenue management for online advertising using advertising networks, Technical report, Working Paper.

Furió, D. and Pardo, Á. (2012), 'Partisan politics theory and stock market performance: Evidence for spain', *Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad* **41**(155), 371–392.

Füss, R. and Bechtel, M. M. (2008), 'Partisan politics and stock market performance: The effect of expected government partisanship on stock returns in the 2002 german federal election', *Public Choice* **135**(3-4), 131–150.

Gallego, G. and Van Ryzin, G. (1994), 'Optimal dynamic pricing of inventories with stochastic demand over finite horizons', *Management science* **40**(8), 999–1020.

Gemmill, G. (1992), 'Political risk and market efficiency: tests based in british stock and options markets in the 1987 election', *Journal of Banking & Finance* **16**(1), 211–231.

Geraghty, M. K. and Johnson, E. (1997), 'Revenue management saves national car rental', *Interfaces* **27**(1), 107–127.

Godfrey, L. G. (1978), 'Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables', *Econometrica: Journal* *of the Econometric Society* pp. 1293–1301.

Gonzalo, J. and Pitarakis, J.-y. (2002), 'Lag length estimation in large dimensional systems', *Journal of Time Series Analysis* **23**(4), 401–423.

Google (2007), 'Official google blog: Latest content id tool for youtube', https://goo.gl/7hGGgG. (Accessed on 08/28/2017).

Google (2008), 'Making money on youtube with content id', https://goo.gl/ZMxkeU. (Accessed on 08/28/2017).

Google (2016*a*), 'The latest video trends: Where your audience is watching', https://goo.gl/kiCBgU. (Accessed on 09/16/2017).

Google (2016*b*), 'The latest video trends: Where your audience is watching', https://goo.gl/1sgtbX. (Accessed on 09/16/2017).

Google (2016*c*), 'Shopping-infographic-1200', https://goo.gl/WGHMQK. (Accessed on 09/16/2017).

Google (2016*d*), 'Top four content categories on youtube'.
URL: *https://www.thinkwithgoogle.com/data/top-content-categories-youtube/*

Gordon, M. J. (1962), *The investment, financing, and valuation of the corporation*, RD Irwin Homewood, IL.

Granger, C. W. (1969), 'Investigating causal relations by econometric models and cross-spectral methods', *Econometrica: Journal of the Econometric Society* pp. 424–438.

Greene, W. H. (2008), *Econometric analysis, 6th ed*, Prentice Hall.

Gujarati, D. N. and Porter, D. C. (2003), 'Basic econometrics. 4th', *New York: McGraw-Hill* .

Gujarati, D. N. and Porter, D. C. (2009), 'Essentials of econometrics'.

Gümüş, B. F. and Zeren, F. (2014), 'Analyzing the efficient market hypothesis with the fourier unit root tests: Evidence from g-20 countries', *Ekonomski horizonti* **16**(3), 225–237.

Hacker, R. S. and Hatemi-J, A. (2006), 'Tests for causality between integrated variables using asymptotic and bootstrap distributions: theory and application', *Applied Economics* **38**(13), 1489–1500.

Halderman, J. A. (2002), Evaluating new copy-prevention techniques for audio cds, *in* 'ACM Workshop on Digital Rights Management', Springer, pp. 101–117.

Hamedy, S. (2016), 'Youtube has paid $2 billion to rights holders through content id', https://mashable.com/2016/07/13/youtube-content-id-piracy-update/#buN17YQ9TSqA. (Accessed on 09/16/2019).

Hannan, E. J. and Quinn, B. G. (1979), 'The Determination of the Order of an Autoregression', *Source Journal of the Royal Statistical Society. Series B (Methodological) Journal of the Royal Statistical Society. Series BMethodological) J. R. Statist. Soc. B* **41**(2), 190–195.

Harvey, C. R. (1989), 'Forecasts of economic growth from the bond and stock markets', *Financial Analysts Journal* **45**(5), 38–45.

Hassanabadi, A. (2011), 'Viacom v. youtube: All eyes blind–the limits of the dmca in a web 2.0 world', *Berkeley technology law journal* .

Hassani, H., Zhigljavsky, A., Patterson, K. and Soofi, A. (2010), 'A comprehensive causality test based on the singular spectrum analysis', *Causality in Science* pp. 379–406.

Hatemi-J, A. and S. Hacker, R. (2009), 'Can the lr test be helpful in choosing the optimal lag order in the var model when information criteria suggest different lag orders?', *Applied Economics* **41**(9), 1121–1125.

Hausman, J. A. (1978), 'Specification tests in econometrics', *Econometrica: Journal of the econometric society* pp. 1251–1271.

Herron, M. C., Lavin, J., Cram, D. and Silver, J. (1999), 'Measurement of political effects in the United States economy: a study of the 1992 presidential election', *Economics and Politics* **44**.

Hibbs, D. (1977*a*), 'Political Parties and Macroeconomic Policy', *American Political Science Review* **71**(4), 1467–1487.

Hibbs, D. (1987), 'The american political economy: Electoral policy and macroeconomics in contemporary america', *Cambridge, MA: Harvard University.[Links]* .

Hibbs, D. A. (1977*b*), 'Political parties and macroeconomic policy', *American political science review* **71**(4), 1467–1487.

Hibbs, D., Carlsen, F. and Pedersen, E. (1996), 'Electoral uncertainty and partisan output cycles', *Research Training Group on Positive Political Economy, Harvard University* .

Hiemstra, C. and Jones, J. D. (1994), 'Testing for linear and nonlinear granger causality in the stock price-volume relation', *The Journal of Finance* **49**(5), 1639–1664.

Ho, T.-H. and Chen, K.-Y. (2007), 'New product blockbusters: The magic and science of prediction markets', *California Management Review* **50**(1), 144–158.

Hobijn, B., Franses, P. H. and Ooms, M. (2004), 'Generalizations of the KPSS-test for stationarity', *Statistica Neerlandica* **58**(4), 483–502.

Holland, B. and Sam, C. (2019), 'A $600 billion bill: Counting the global cost of the u.s.-china trade war'.
URL: *https://www.bloomberg.com/graphics/2019-us-china-trade-war-economic-fallout/*

Homaifar, G., Randolph, W. L., Helms, B. P. and Haddad, M. (1988), 'American presidential elections and returns of defence industry stocks', *Applied Economics* **20**(7), 985–993.

Hormann, K. C. (2009), 'The death of the dmca-how viacom v. youtube may define the future of digital content', *Hous. L. Rev.* **46**, 1345.

Hotten, R. (2019), 'Trump's threat of mexico trade war sends markets lower'.
URL: *https://www.bbc.com/news/business-48451342*

Howe, J. S. (1986), 'Evidence on stock market overreaction', *Financial Analysts Journal* **42**(4), 74–77.

Hsiao, C. (1985), 'Benefits and limitations of panel data', *Econometric Reviews* **4**(1), 121–174.

Hsiao, C. (2003), *Analysis of panel data*, Cambridge university press.

Hsiao, C. (2007), 'Panel data analysis - advantages and challenges', *Sociedad de Estad* *ıstica e Investigaci on Operativa* **00**(0), 1–63.

Hsiao, C., Mountain, D. C. and Illman, K. H. (1995), 'A bayesian integration of end-use metering and conditional-demand analysis', *Journal of Business & Economic Statistics* **13**(3), 315–326.

Hu, S., Cao, Y., Zhang, J., Kong, W., Yang, K., Zhang, Y. and Li, X. (2012), 'Granger causality's shortcomings and new causality measure', *Cognitive Neurodynamics* **6**, 33–42.

Huang, R. D. (1985), 'Common stock returns and presidential elections', *Financial Analysts Journal* **41**(2), 58–61.

Hubert, M. and Vandervieren, E. (2008), 'An adjusted boxplot for skewed distributions', *Computational statistics & data analysis* **52**(12), 5186–5201.

HubSpot (2016), 'The future of content marketing: How people are changing the way they read, interact, and engage with content', https://goo.gl/J17BYL. (Accessed on 09/16/2017).

HubSpot (2017), 'State of inbound - marketing & sales data', https://goo.gl/MRrHFW. (Accessed on 09/16/2017).

Hudson, R., Keasey, K. and Dempsey, M. (1998), 'Share prices under tory and labour governments in the uk since 1945', *Applied Financial Economics* **8**(4), 389–400.

Hung, J.-C., Jiang, S.-J. and Chiu, C.-L. (2007), 'Jump risk of presidential election: evidence from taiwan stock and foreign exchange markets', *Applied Economics* **39**(17), 2231–2240.

Hunt, K. (2007), 'Copyright and youtube: Pirate's playground or fair use forum', *Mich. Telecomm. & Tech. L. Rev.* **14**, 197.

Hyndman, R. J. and Athanasopoulos, G. (2014), *Forecasting: principles and practice*, OTexts.

IABCanada (2018), 'Canadian internet advertising revenue survey', http://iabcanada.com/content/uploads/2019/02/IABCanada_2017-Revenue-Exec-Report-_Dec1018.pdf. (Accessed on 10/18/2019).

IBM (2015).

URL: *https://ibm.co/2pPzxMA*

ICCBelgium (2017), 'Global impacts of counterfeiting and piracy', https://goo.gl/KUAjRQ. (Accessed on 10/03/2017).

IFPI (2015), 'Digital music report 2015: Charting the path to sustainable growth 6 (2015)'. (Accessed on 08/26/2019).

IFPI (2016), 'Global music report 2016: State of the industry', https://goo.gl/hk77MK. (Accessed on 08/26/2017).

IMF (2018), 'Imf', https://www.imf.org/external/datamapper/NGDPD@WEO/WEOWORLD/USA. (Accessed on 09/27/2018).

Iqbal, M. (2019), 'Youtube revenue and usage statistics (2019)'.

URL: *https://bit.ly/2pGG4t5*

Jacobson, G. C. (1990), 'Does the Economy Matter in Midterm Elections?'.

Jensen, M. C. (1978), 'Some anomalous evidence regarding market efficiency', *Journal of financial economics* **6**(2-3), 95–101.

Johansen, S. and Juselius, K. (1990), 'Maximum likelihood estimation and inference on cointegration with applications to the demand for money.', *Oxford Bulletin of Economics and Statistics* **52**(2), 169–210.

Jones, S. D. (2019), 'Dow jones global indices methodology - s&p dow jones indices', https://bit.ly/2YZfZpi. (Accessed on 08/21/2019).

Jung, A. (2014), 'User-generated parody videos on youtube: Do they have a successful fair use defense under copyright law?', *Law School Student Scholarship* .

Kallas, P. (2018), '8 powerful reasons you need to use video marketing [trends]'.

URL: *https://www.dreamgrow.com/8-reasons-why-your-business-should-use-video-marketing/*

Karnizova, L. and Li, J. C. (2014), 'Economic policy uncertainty, financial markets and probability of us recessions', *Economics Letters* **125**(2), 261–265.

Kelly, B., Pástor, L. and Veronesi, P. (2016), 'The price of political uncertainty: Theory and evidence from the option market', *The Journal of Finance* **71**(5), 2417–2480.

Kim, E. C. (2007), 'Youtube: Testing the safe harbors of digital copyright law', *S. Cal. Interdisc. LJ* **17**, 139.

Kim, H. Y. and Mei, J. P. (2001), 'What makes the stock market jump? an analysis of political risk on hong kong stock returns', *Journal of International Money and Finance* **20**(7), 1003–1016.

Kim, J. H. and Shamsuddin, A. (2008), 'Are asian stock markets efficient? evidence from new multiple variance ratio tests', *Journal of Empirical Finance* **15**(3), 518–532.

Knight, B. G. (2004), 'Are Policy Platforms Capitalized into Equity Prices? Evidence from the Bush/Gore2000 Presidential Election', *Journal of Public Economics* **NBER Worki**(4), 751–773.

Kruskal, W. H. and Wallis, W. A. (1952), 'Use of ranks in one-criterion variance analysis', *Journal of the American statistical Association* **47**(260), 583–621.

Kumar, P. (2008), 'Locating the boundary between fair use and copyright infringement: the viacom–youtube dispute', *Journal of Intellectual Property Law & Practice* **3**(12), 775–778.

Kumar, S. and Sethi, S. P. (2009), 'Dynamic pricing and advertising for web content providers', *European Journal of Operational Research* **197**(3), 924–944.

Kwiatkowski, D., Phillips, P. C., Schmidt, P. and Shin, Y. (1992), 'Testing the null hypothesis of stationarity against the alternative of a unit root', *Journal of Econometrics* **54**(1-3), 159–178.

Kwok, S. H. (2002), 'Digital rights management for the online music business', *ACM Sigecom exchanges* **3**(3), 17–24.

Lai, T. L. and Robbins, H. (1985), 'Asymptotically efficient adaptive allocation rules', *Advances in applied mathematics* **6**(1), 4–22.

Lamb, R. P., Ma, K., Pace, R. D. and Kennedy, W. F. (1997), 'The congressional calendar and stock market performance', *Financial Services Review* **6**(1), 19–25.

Lin, C. and Wang, Y. (2004), 'Effects of the legislative sessions on the stock market: Evidence from taiwan', *Indian Journal of Economics* **336**, 1–16.

Lin, M., Ke, X. and Whinston, A. B. (2012), 'Vertical differentiation and a comparison of online advertising models', *Journal of Management Information Systems* **29**(1), 195–236.

Lin, M., Lucas, H. C. and Shmueli, G. (2011), 'Too big to fail: Larger samples and false discoveries', *Robert H. Smith School Research Paper No. RHS* pp. 06–068.

Littlewood, K. (1972), 'Forecasting and control of passenger bookings', *Airline Group International Federation of Operational Research Societies Proceedings, 1972* **12**, 95–117.

Liu, L. and Zhang, T. (2015), 'Economic policy uncertainty and stock market volatility', *Finance Research Letters* **15**, 99–105.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer.

Lyon, G. E. (2001), 'The internet marketplace and digital rights management', *National Institute for Standards and Technology* .

MacKinnon, J. G. and White, H. (1985), 'Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties', *Journal of econometrics* **29**(3), 305–325.

Maddala, G. and Lahiri, K. (2001), *Introduction to Econometrics*, West Sussex, John Wiley &Sons Ltd.

Mahadeva, L. and Robinson, P. (2004), 'Unit Root Testing to Help Model Building', *Handbook in Central Banking* (22), 1–53.

Makovskỳ, P. (2014), 'Modern approaches to efficient market hypothesis of forex–the central european case', *Procedia Economics and Finance* **14**, 397–406.

Malkiel, B. (1992), 'Efficient market hypothesis'.

Mann, H. B. and Whitney, D. R. (1947), 'On a test of whether one of two random variables is stochastically larger than the other', *The annals of mathematical statistics* pp. 50–60.

Marvin, G. (2017), 'Report: Google earns 78% of $36.7b us search ad revenues, soon to be 80%', https://goo.gl/ER7gD2. (Accessed on 10/18/2017).

Masnick, M. (2013), 'Youtube wins yet another complete victory over viacom', https://goo.gl/uq5n9j. (Accessed on 08/28/2017).

McKenzie, J. (2009), 'Illegal music downloading and its impact on legitimate sales: Australian empirical evidence', *Australian Economic Papers* **48**(4), 296–307.

Mundlak, Y. (1978), 'On the pooling of time series and cross section data', *Econometrica: journal of the Econometric Society* pp. 69–85.

Musau, V. M., Waititu, A. G. and Wanjoya, A. K. (2015), 'Modeling panel data: Comparison of gls estimation and robust covariance matrix estimation', *American Journal of Theoretical and Applied Statistics* **4**(3), 185–191.

Musk, E. (2019), 'All our patent are belong to you'.
URL: *https://www.tesla.com/blog/all-our-patent-are-belong-you*

Nair, S. K. and Bapna, R. (2001), 'An application of yield management for internet service providers', *Naval Research Logistics (NRL)* **48**(5), 348–362.

Newey, W. K. and West, K. D. (1986), 'A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix'.

Ng, B. Y. S. and Perron, P. (2001), 'Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power', *Econometrica* **69**(6), 1519–1554.

Nickles, M. (2004), 'Presidential elections and stock market cycles', *Graziadio Business Review* **7**(3).

Niederhoffer, V., Gibbs, S. and Bullock, J. (1970), 'Presidential elections and the stock market', *Financial Analysts Journal* pp. 111–113.

Nordhaus, W. D. (1975), 'The political business cycle', *The review of economic studies* **42**(2), 169–190.

Oberholzer-Gee, F. and Strumpf, K. (2007), 'The effect of file sharing on record sales: An empirical analysis', *Journal of political economy* **115**(1), 1–42.

O'Brien, D. (2008), 'Copyright challenges for user generated intermediaries: Viacom v youtube and google'.

Oehler, A., Walker, T. J. and Wendt, S. (2013), 'Effects of election results on stock price performance: evidence from 1980 to 2008', *Managerial Finance* **39**(8), 714–736.

Office, U. C. (1998), 'The digital millennium copyright act of 1998', https://www.copyright.gov/legislation/dmca.pdf. (Accessed on 08/26/2017).

Omnicore (2019), 'Youtube by the numbers: Stats, demographics & fun facts'.
URL: *https://www.omnicoreagency.com/youtube-statistics/*

Pandey, S., Agarwal, D., Chakrabarti, D. and Josifovski, V. (2007), Bandits for taxonomies: A model-based approach, *in* 'Proceedings of the 2007 SIAM International Conference on Data Mining', SIAM, pp. 216–227.

Pantzalis, C., Stangeland, D. A. and Turtle, H. J. (2000), 'Political elections and the resolution of uncertainty: the international evidence', *Journal of banking & finance* **24**(10), 1575–1604.

Papies, D., Eggers, F. and Wlömert, N. (2011), 'Music for free? how free ad-funded downloads affect consumer choice', *Journal of the Academy of Marketing Science* **39**(5), 777–794.

Pástor, L. and Veronesi, P. (2013), 'Political uncertainty and risk premia', *Journal of Financial Economics* **110**(3), 520–545.

Patten, L. B. (2007), 'From safe harbor to choppy waters: Youtube, the digital millennium copyright act, and a much needed change of course', *Vand. J. Ent. & Tech. L.* **10**, 179.

Persson, T., Roland, G. and Tabellini, G. (2003), How do electoral rules shape party structures, government coalitions, and economic policies?, Technical report, National Bureau of Economic Research.

Petersen, M. A. (2009), 'Estimating standard errors in finance panel data sets: Comparing approaches', *The Review of Financial Studies* **22**(1), 435–480.

Phillips, P. C. and Durlauf, S. N. (1986), 'Multiple time series regression with integrated processes', *The Review of Economic Studies* **53**(4), 473–495.

Phillips, P. and Perron, P. (1988), 'Testing for a Unit Root in Time Series Regressions', *Biometrika* **75**(2), 335–346.

Pitkänen, O. and Välimäki, M. (2000), 'Towards a digital rights management framework'.

Radovanovic, A. and Zeevi, A. (2010), 'Revenue maximization through'smart'inventory management in reservation-based online advertising', *SIGMETRICS Performance Evaluation Review* **38**(2), 33–35.

Rambaldi, A. N. and Doran, H. E. (1996), *Testing for granger non-casuality in cointegrated systems made easy*, Department of Econometrics, University of New England.

Reed, W. R. and Ye, H. (2011), 'Which panel data estimator should i use?', *Applied economics* **43**(8), 985–1000.

Roberts, B. E. (1990*a*), 'Political institutions, policy expectations, and the 1980 election: a financial market perspective', *American Journal of Political Science* pp. 289–310.

Roberts, B. E. (1990*b*), 'Political Institutions, Policy Expectations, and the 1980 Election: A Financial Market Perspective', *American Journal of Political Science* **34**(2), 289–310.

Rogoff, K. (1990), 'Equilibrium political budget cycles', *American Economic Review* **80**(1), 21–36.

Rosenblatt, B., Trippe, B., Mooney, S. et al. (2002), 'Digital rights management', *New York* .

Rothschild, D. (2009), 'Forecasting elections: Comparing prediction markets, polls, and their biases', *Public Opinion Quarterly* **73**(5), 895–916.

Rothstein, M. (1971), 'An airline overbooking model', *Transportation Science* **5**(2), 180–192.

Rothstein, M. (1974), 'Hotel overbooking as a markovian sequential decision process', *Decision Sciences* **5**(3), 389–404.

Roubini, N. and Sachs, J. (1989), 'Government spending and budget deficits in the industrial countries', *Economic policy* **4**(8), 99–132.

Sandvine (2016), '2016 - global internet phenomena - latin america & north america', https://goo.gl/4fkq4c. (Accessed on 09/16/2017).

Santa-Clara, P. and Valkanov, R. (2003), 'The presidential puzzle: Political cycles and the stock market', *The Journal of Finance* **58**(5), 1841–1872.

Satterthwaite, F. E. (1946), 'An approximate distribution of estimates of variance components', *Biometrics bulletin* **2**(6), 110–114.

Sauré, D. and Zeevi, A. (2013), 'Optimal dynamic assortment planning with demand learning', *Manufacturing & Service Operations Management* **15**(3), 387–404.

Schwarz, G. (1978), 'Estimating the Dimension of a Model', *Annals of Statistics* **6**(2), 461–464.

Schwert, G. W. (2002), 'Tests for unit roots: A monte carlo investigation', *Journal of Business & Economic Statistics* **20**(1), 5–17.

Scott Hacker, R. and Hatemi-J, A. (2008), 'Optimal lag-length choice in stable and unstable VAR models under situations of homoscedasticity and ARCH', *Journal of Applied Statistics* **35**(6), 601–615.

Scott Long, J. and Ervin, L. (1998), 'Correcting for heteroscedasticity with heteroscedasticity consistent standard errors in the linear regression model: Small sample considerations', *Indiana University Bloomington, IN* **47405**.

Seth, A. (2007), 'Granger causality', *Scholarpedia* **2**(7), 1667.

Shapiro, S. S. and Wilk, M. B. (1965), 'An analysis of variance test for normality (complete samples)', *Biometrika* **52**(3/4), 591–611.

Sims, C. A. (1972), 'Money, income, and causality', *The American economic review* **62**(4), 540–552.

Siwek, S. E. (2007), 'The true cost of sound recording piracy to the us economy'.

Smith, A. K. (2016), 'How presidential elections affect the stock market'.
URL: *https://www.kiplinger.com/article/investing/T043-C008-S003-how-presidential-elections-affect-the-stock-market.html*

Smith, B. C., Leimkuhler, J. F. and Darrow, R. M. (1992), 'Yield management at american airlines', *interfaces* **22**(1), 8–31.

Snowberg, E., Wolfers, J. and Zitzewitz, E. (2007), 'Party Influence in Congress and the Economy', *Quarterly Journal of Political Science* **2**(3), 277–286.

Solomon, L. (2015), 'Fair users or content abusers: The automatic flagging of non-infringing videos by content id on youtube', *Hofstra L. Rev.* **44**, 237.

Spangler, T. (2014), 'Why the viacom-youtube case wasn't a giant waste of time'.
URL: *https://variety.com/2014/biz/news/why-the-viacom-youtube-case-wasnt-a-giant-waste-of-time-1201137664/*

Spangler, T. (2020), 'Alphabet reports youtube ad revenue for first time, video service generated $15.1 billion in 2019'.
URL: *https://variety.com/2020/digital/news/alphabet-youtube-ad-revenue-first-time-15-billion-2019-1203491155/!*

statisticBrain (2017), 'Youtube company statistics — statistic brain', https://goo.gl/t1F6Za. (Accessed on 08/26/2017).

Steeley, J. M. (2003), 'Making political capital: the behaviour of the uk capital markets during election'97', *Applied financial economics* **13**(2), 85–95.

Stern, D. I. and Enflo, K. (2013), 'Causality between energy and output in the long-run', *Energy economics* **39**, 135–146.

Stock, J. H. and Watson, M. W. (2008), 'Heteroskedasticity-robust standard errors for fixed effects panel data regression', *Econometrica* **76**(1), 155–174.

Stovall, R. H. (1992), 'Forecasting stock market performance via the presidential cycle', *Financial Analysts Journal* pp. 5–8.

Sturm, R. R. (2009), 'The 'other'january effect and the presidential election cycle', *Applied Financial Economics* **19**(17), 1355–1363.

Subrahmanyan, S. and Shoemaker, R. (1996), 'Developing optimal pricing and inventory policies for retailers who face uncertain demand', *Journal of Retailing* **72**(1), 7–30.

Talluri, K. and Van Ryzin, G. (2004), 'Revenue management under a general discrete choice model of consumer behavior', *Management Science* **50**(1), 15–33.

Tassi, P. (2013), 'Youtube unleashes strange storm of copyright claims on video game content producers', https://goo.gl/7Gsduj. (Accessed on 08/28/2017).

Toda, H. Y. and Yamamoto, T. (1995), 'Statistical inference in vector autoregressions with possibly integrated processes', *Journal of econometrics* **66**(1), 225–250.

Tokic, D. (2003), 'Emerging markets before the 1997 asia pacific financial crisis', *Asia Pacific Business Review* **9**(3), 105–115.

TubeMogul (2017), 'Digital advertising formats', https://goo.gl/V2N167. (Accessed on 10/19/2017).

Tuck, C. E. and Hon, L. Y. (2008), 'A note on the general elections and long memory: evidence from the london stock exchange', *Applied Financial Economics Letters* **4**(5), 331–335.

Urrutia, J. L. (1995), 'Tests of random walk and market efficiency for latin american emerging equity markets', *Journal of financial research* **18**(3), 299–309.

Vargas, P. T., Duff, B. R. and Faber, R. J. (2017), 'A practical guide to experimental advertising research', *Journal of Advertising* **46**(1), 101–114.

Vernik, D. A., Purohit, D. and Desai, P. S. (2011), 'Music downloads and the flip side of digital rights management', *Marketing Science* **30**(6), 1011–1027.

VerSteeg, R. (2007), 'Viacom v. youtube: Preliminary observations', *NCJL & Tech.* **9**, 43.

videonitch (2017), '36 mind blowing youtube facts', https://bit.ly/32nRDSX. (Accessed on 10/03/2017).

vidYard (2014), 'Marketers who use video outperform their peers', https://goo.gl/tufa22. (Accessed on 09/16/2017).

Volkerink, B. and De Haan, J. (2001), 'Fragmented government effects on fiscal policy: New evidence', *Public choice* **109**(3-4), 221–242.

Wang, Y., Chen, C. R. and Huang, Y. S. (2014), 'Economic policy uncertainty and corporate investment: Evidence from china', *Pacific-Basin Finance Journal* **26**, 227–243.

Wang, Y.-H. and Lin, C.-T. (2008), 'Empirical analysis of political uncertainty on taiex stock market', *Applied Economics Letters* **15**(7), 545–550.

White, H. et al. (1980), 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *econometrica* **48**(4), 817–838.

Wikström, P. (2013), *The music industry: Music in the cloud*, Polity.

Wisniewski, T. P. (2016), 'Is there a link between politics and stock returns? A literature survey', *International Review of Financial Analysis* **47**, 15–23.
URL: *http://dx.doi.org/10.1016/j.irfa.2016.06.015*

Wooldridge, J. M. (2003), 'Cluster-sample methods in applied econometrics', *American Economic Review* **93**(2), 133–138.

Wooldridge, J. M. (2010), *Econometric analysis of cross section and panel data*, MIT press.

Yang, J., Vee, E., Vassilvitskii, S., Tomlin, J., Shanmugasundaram, J., Anastasakos, T. and Kennedy, O. (2010), 'Inventory allocation for online graphical display advertising', *arXiv preprint arXiv:1008.3551* .

YouTube (2017*a*), 'Press - youtube', https://goo.gl/cJsXzR. (Accessed on 09/16/2017).

YouTube (2017*b*), 'Understand youtube rights management', https://goo.gl/jTvBAc. (Accessed on 08/28/2017).

YouTube (2017*c*), 'Using content id - youtube help', https://goo.gl/b7vW3E. (Accessed on 08/28/2017).

YouTube (2019), 'Youtube for press', https://goo.gl/TeQJVb. (Accessed on 08/28/2019).

Zabinsky, Z. B. (2010), 'Random search algorithms', *Wiley Encyclopedia of Operations Research and Management Science* .

Zapata, H. O. and Rambaldi, A. N. (1997), 'Monte Carlo Evidence on Cointegration and Causation', *Oxford Bulletin of Economics and Statistics* **59**(2), 285–298.

Zivot, E. and Wang, J. (2006), 'Vector autoregressive models for multivariate time series', *Modeling Financial Time Series with S-Plus®* pp. 385–429.

APPENDICES

Appendix A

Appendix A

A.1 Fixed effect test results with sub-sampled datasets

To make sure our large sample sizes do not affect our inferences, we randomly sub-sampled both datasets 100 times with different sample sizes for our statistical tests. Table A.1 and A.2 the result of the fixed-effect models over 100 sub-sampled TC and MC datasets, respectively. In both tables, the "sample size" column shows the number of videos that we included in the sub-sampled datasets. In each subsampling iterating, the number of videos decreased (e.g., Table A.1 shows that the last subsample with id equal 100, we have 270 videos in the sub-sampled panel data).

The results of sub-sampled fixed effect tests, presented in Tables A.1 and A.2 ,show that the p-values' significance is the same with all sample sizes. All p-values presented in Table A.1 are greater than five percent (i.e., the coefficient of TC models are not significant). In Table A.2 all p-values are near zero, which means the the coefficient of MC models are significant.

Table A.1: Tracker Channel dataset - 100 sub-samples test results

id	sample size	std-err	t-stat	p-value	id	sample size	std-err	t-stat	p-value
1	3240	0.0027	0.81	0.42	51	1740	0.0037	0.62	0.53
2	3210	0.0027	0.87	0.39	52	1710	0.0039	1.30	0.19
3	3180	0.0028	0.86	0.39	53	1680	0.0038	1.41	0.16
4	3150	0.0028	0.66	0.51	54	1650	0.0038	0.44	0.66
5	3120	0.0028	0.16	0.87	55	1620	0.0039	0.32	0.75
6	3090	0.0028	0.54	0.59	56	1590	0.0040	0.00	1.00
7	3060	0.0028	0.77	0.44	57	1560	0.0038	0.65	0.52
8	3030	0.0029	0.66	0.51	58	1530	0.0040	0.21	0.83
9	3000	0.0028	1.03	0.30	59	1500	0.0041	0.86	0.39
10	2970	0.0029	1.17	0.24	60	1470	0.0043	1.15	0.25
11	2940	0.0029	0.94	0.35	61	1440	0.0041	-0.31	0.75
12	2910	0.0029	1.00	0.32	62	1410	0.0042	-0.99	0.32
13	2880	0.0029	0.13	0.90	63	1380	0.0044	1.02	0.31
14	2850	0.0029	0.66	0.51	64	1350	0.0045	1.58	0.11
15	2820	0.0030	0.83	0.40	65	1320	0.0043	1.12	0.26
16	2790	0.0030	0.84	0.40	66	1290	0.0043	0.64	0.52
17	2760	0.0030	1.09	0.28	67	1260	0.0042	0.06	0.95
18	2730	0.0030	1.35	0.18	68	1230	0.0045	0.80	0.43
19	2700	0.0030	1.09	0.28	69	1200	0.0046	1.28	0.20
20	2670	0.0030	0.59	0.55	70	1170	0.0044	1.07	0.29
21	2640	0.0030	0.23	0.81	71	1140	0.0047	0.59	0.55
22	2610	0.0030	0.46	0.65	72	1110	0.0047	-0.20	0.84
23	2580	0.0031	1.03	0.30	73	1080	0.0046	1.24	0.21
24	2550	0.0031	1.03	0.30	74	1050	0.0049	0.26	0.80
25	2520	0.0031	0.15	0.88	75	1020	0.0049	-0.36	0.72
26	2490	0.0032	0.73	0.47	76	990	0.0050	-0.39	0.70
27	2460	0.0032	0.83	0.41	77	960	0.0051	0.55	0.58
28	2430	0.0031	-0.01	0.99	78	930	0.0049	-0.71	0.48
29	2400	0.0033	0.56	0.57	79	900	0.0051	0.03	0.98
30	2370	0.0032	1.07	0.28	80	870	0.0054	-0.29	0.78
31	2340	0.0032	0.03	0.98	81	840	0.0051	2.12	0.03
32	2310	0.0032	0.47	0.64	82	810	0.0055	1.22	0.22
33	2280	0.0033	-0.20	0.84	83	780	0.0057	0.46	0.65
34	2250	0.0033	0.89	0.37	84	750	0.0052	0.68	0.50
35	2220	0.0033	0.97	0.33	85	720	0.0061	0.32	0.75
36	2190	0.0034	0.18	0.85	86	690	0.0061	-0.42	0.68
37	2160	0.0033	0.36	0.72	87	660	0.0058	0.11	0.91
38	2130	0.0035	0.95	0.34	88	630	0.0065	-0.50	0.62
39	2100	0.0035	0.13	0.89	89	600	0.0062	0.76	0.45
40	2070	0.0035	1.22	0.22	90	570	0.0069	1.31	0.19
41	2040	0.0034	0.40	0.69	91	540	0.0066	1.20	0.23
42	2010	0.0034	1.03	0.30	92	510	0.0069	0.76	0.45
43	1980	0.0035	-0.04	0.97	93	480	0.0071	0.20	0.84
44	1950	0.0035	-0.03	0.98	94	450	0.0070	-0.62	0.54
45	1920	0.0036	0.21	0.83	95	420	0.0078	-0.11	0.92
46	1890	0.0036	1.08	0.28	96	390	0.0081	-0.16	0.87
47	1860	0.0037	0.92	0.36	97	360	0.0075	-0.69	0.49
48	1830	0.0037	1.15	0.25	98	330	0.0093	1.45	0.15
49	1800	0.0036	-0.21	0.83	99	300	0.0096	0.97	0.33
50	1770	0.0038	0.21	0.84	100	270	0.0086	0.76	0.45

Table A.2: Montizer Channel dataset - 100 sub-samples test results

id	sample size	std-err	t-stat	p-value	id	sample size	std-err	t-stat	p-value
1	18537	0.0002	15.03	0.00	51	17037	0.0002	14.77	0.00
2	18507	0.0002	15.03	0.00	52	17007	0.0002	14.28	0.00
3	18477	0.0002	15.01	0.00	53	16977	0.0002	14.77	0.00
4	18447	0.0002	15.01	0.00	54	16947	0.0002	14.02	0.00
5	18417	0.0002	15.02	0.00	55	16917	0.0002	14.10	0.00
6	18387	0.0002	15.12	0.00	56	16887	0.0002	14.28	0.00
7	18357	0.0002	14.77	0.00	57	16857	0.0002	14.25	0.00
8	18327	0.0002	14.97	0.00	58	16827	0.0002	14.63	0.00
9	18297	0.0002	15.08	0.00	59	16797	0.0002	14.25	0.00
10	18267	0.0002	14.88	0.00	60	16767	0.0002	14.46	0.00
11	18237	0.0002	14.68	0.00	61	16737	0.0002	14.60	0.00
12	18207	0.0002	14.85	0.00	62	16707	0.0002	14.42	0.00
13	18177	0.0002	14.82	0.00	63	16677	0.0002	14.33	0.00
14	18147	0.0002	14.75	0.00	64	16647	0.0002	14.23	0.00
15	18117	0.0002	14.68	0.00	65	16617	0.0002	14.62	0.00
16	18087	0.0002	14.85	0.00	66	16587	0.0002	14.36	0.00
17	18057	0.0002	14.87	0.00	67	16557	0.0002	14.46	0.00
18	18027	0.0002	14.71	0.00	68	16527	0.0002	14.24	0.00
19	17997	0.0002	14.80	0.00	69	16497	0.0002	14.43	0.00
20	17967	0.0002	14.86	0.00	70	16467	0.0002	13.70	0.00
21	17937	0.0002	14.46	0.00	71	16437	0.0002	14.50	0.00
22	17907	0.0002	14.82	0.00	72	16407	0.0002	14.23	0.00
23	17877	0.0002	15.38	0.00	73	16377	0.0002	13.83	0.00
24	17847	0.0002	14.62	0.00	74	16347	0.0002	13.84	0.00
25	17817	0.0002	14.49	0.00	75	16317	0.0002	13.70	0.00
26	17787	0.0002	14.50	0.00	76	16287	0.0002	14.31	0.00
27	17757	0.0002	14.87	0.00	77	16257	0.0002	14.65	0.00
28	17727	0.0002	14.47	0.00	78	16227	0.0002	14.52	0.00
29	17697	0.0002	14.87	0.00	79	16197	0.0002	13.65	0.00
30	17667	0.0002	14.90	0.00	80	16167	0.0002	14.15	0.00
31	17637	0.0002	14.69	0.00	81	16137	0.0002	14.06	0.00
32	17607	0.0002	14.76	0.00	82	16107	0.0002	14.19	0.00
33	17577	0.0002	14.58	0.00	83	16077	0.0002	13.66	0.00
34	17547	0.0002	14.99	0.00	84	16047	0.0002	13.90	0.00
35	17517	0.0002	14.86	0.00	85	16017	0.0002	14.08	0.00
36	17487	0.0002	14.23	0.00	86	15987	0.0002	13.85	0.00
37	17457	0.0002	14.47	0.00	87	15957	0.0002	14.04	0.00
38	17427	0.0002	14.67	0.00	88	15927	0.0002	14.01	0.00
39	17397	0.0002	14.49	0.00	89	15897	0.0002	13.57	0.00
40	17367	0.0002	14.91	0.00	90	15867	0.0002	13.66	0.00
41	17337	0.0002	14.60	0.00	91	15837	0.0002	14.03	0.00
42	17307	0.0002	14.55	0.00	92	15807	0.0002	13.86	0.00
43	17277	0.0002	14.92	0.00	93	15777	0.0002	13.80	0.00
44	17247	0.0002	14.56	0.00	94	15747	0.0002	13.60	0.00
45	17217	0.0002	14.28	0.00	95	15717	0.0002	13.45	0.00
46	17187	0.0002	14.18	0.00	96	15687	0.0002	13.47	0.00
47	17157	0.0002	14.29	0.00	97	15657	0.0002	13.35	0.00
48	17127	0.0002	14.54	0.00	98	15627	0.0002	13.85	0.00
49	17097	0.0002	14.32	0.00	99	15597	0.0002	14.38	0.00
50	17067	0.0002	14.81	0.00	100	15567	0.0002	13.77	0.00

Appendix B

Appendix B

B.1 YouTube Traffic Sources Definition

- **ADVERTISING** - The viewer was referred to the video by an advertisement.
- **ANNOTATION** - Viewers reached the video by clicking on an annotation in another video.
- **CAMPAIGN_CARD** – Views originated from claimed, user-uploaded videos that the content owner used to promote the viewed content. This traffic source is only supported for content owner reports.
- **END_SCREEN** – The views were referred from the end screen of another video.
- **EXT_URL** – The video views were referred from a link on another website.
- **NO_LINK_EMBEDDED** – The video was embedded on another website when it was viewed.
- **NO_LINK_OTHER** – YouTube did not identify a referrer for the traffic. This category encompasses direct traffic to a video as well as traffic on mobile apps.
- **NOTIFICATION** – The video views were referred from an email or notification from YouTube.

- **PLAYLIST** – The video views occurred while the video was being played as part of a playlist.
- **PROMOTED** – The video views were referred from an unpaid YouTube promotion, such as the YouTube "Spotlight Videos" page.
- **RELATED_VIDEO** The video views were referred from a related video listing on another video watch page.
- **SUBSCRIBER** – The video views were referred from feeds on the YouTube homepage or from YouTube subscription features.
- **YT_CHANNEL** – The video views occurred on a channel page.
- **YT_OTHER_PAGE** – The video views were referred from a link other than a search result or related video link that appeared on a YouTube page.
- **YT_PLAYLIST_PAGE** – The video views originated from a page that lists all of the videos in a playlist. Note that this traffic source is different from PLAYLIST, which indicates that the views occurred while the video was being played as part of a playlist.
- **YT_SEARCH** – The video views were referred from YouTube search results.

B.2 Wilcoxon pairwise comparison p-values (one-tailed)

Table B.1 and B.2 show the results of Wilcoxon test that compares all treatment groups to one another. Table B.1 has two panels, A and B, related to the first round of the experiment. Panel A displays the Wilcoxon p-values for the post-test dataset, including outliers. This panel has two sub-panels (i.e., A_1 and A_2) that show p-values for both KPIs (i.e., views in A_1 and gross revenue in A_2). For example, in sub-panel A_1, the first column, T_1, shows the p-values of pairwise comparison of treatment group T_1 to the other treatment groups (i.e., T_2 *to* T_8). In this column, the only p-value less than 0.05% is for the pair (T_1, T_5) with the p-value of 0.01. Then, we conclude that the T_1 average views differs from the T_5 average views when we include outliers. Also, the p-value of (T_5, T_6) is 0.03 in Panel B_2. Thus, we conclude that the T_5 average gross revenue differs from the T_6 average gross revenue when we exclude outliers.

Table B.1: Wilcoxon p-values - First round of the experiment

A	A- dataset with outliers													
	A_1- views							A_2- grossRevenue						
	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_1	T_2	T_3	T_4	T_5	T_6	T_7
T_2	1.00							1.00						
T_3	1.00	0.00						1.00	0.00					
T_4	1.00	0.00	0.44					1.00	1.00	1.00				
T_5	0.01	0.00	0.00	0.00				1.00	0.00	0.00	0.00			
T_6	1.00	0.00	0.14	1.00	1.00			1.00	0.00	0.00	0.00	1.00		
T_7	1.00	0.00	1.00	1.00	1.00	1.00		1.00	1.00	1.00	0.00	1.00	1.00	
T_8	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

B	B- dataset without outliers													
	B_1- views							B_2- grossRevenue						
	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_1	T_2	T_3	T_4	T_5	T_6	T_7
T_2	1.00							1.00						
T_3	1.00	0.00						1.00	0.00					
T_4	1.00	0.00	1.00					1.00	1.00	1.00				
T_5	1.00	0.00	0.00	0.00				1.00	0.00	0.00	0.00			
T_6	1.00	0.00	0.00	0.00	1.00			1.00	0.00	0.00	0.00	0.03		
T_7	1.00	0.00	1.00	1.00	1.00	1.00		1.00	1.00	1.00	0.00	1.00	1.00	
T_8	1.00	0.00	0.05	0.50	1.00	1.00	0.03	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Same as Table B.1, Table B.2 shows the Wilcoxon tests' p-values for both KPIs in the second round of the experiment. Table B.2 also has two main panels (i.e. A and B) and each panel has two sub-panels for both KPIs (i.e., A_1 and B_1 for views and A_2 and B_2 for gross revenue).

Table B.2: Wilcoxon p-values - Second round of the experiment

	A- dataset with outliers						B- dataset without outliers					
	A_1- views			A_2- grossRevenue			B_1- views			B_2- grossRevenue		
	T_1	T_2	T_5	T_1	T_2	T_5	T_1	T_2	T_5	T_1	T_2	T_5
T_2	0.00			0.04			0.05			0.14		
T_5	1.00	1.00		1.00	1.00		1.00	1.00		1.00	1.00	
T_6	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00

B.3 Traffic-source datasets sample

Table B.3: YouTube videos Traffic sources

Day	Traffic Source Type	Number of Views	Views Percentage
2019-08-28	PLAYLIST	440	0.49
2019-08-28	RELATED_VIDEO	132	0.15
2019-08-28	YT_SEARCH	101	0.11
2019-08-28	SUBSCRIBER	84	0.09
2019-08-28	YT_PLAYLIST_PAGE	64	0.07
2019-08-28	YT_OTHER_PAGE	38	0.04
2019-08-28	NO_LINK_OTHER	21	0.02
2019-08-28	YT_CHANNEL	9	0.01
2019-08-28	EXT_URL	5	0.01

B.4 Kruskal Wallis tests with sub-sampled datasets

Table B.4: Kruskal Wallis tests with the sub-sampled datasets - First round experiment

	Views		Gross Revenue				Views		Gross Revenue		
Id	χ^2	p-value	χ^2	p-value	sample size	Id	χ^2	p-value	χ^2	p-value	sample size
1	4.01	0.78	20.65	0.00	2410	51	214.35	0.00	1147.05	0.00	122910
2	19.86	0.01	62.27	0.00	4820	52	181.81	0.00	1208.27	0.00	125320
3	34.36	0.00	116.17	0.00	7230	53	252.48	0.00	1310.03	0.00	127730
4	20.58	0.00	102.85	0.00	9640	54	235.93	0.00	1251.52	0.00	130140
5	13.79	0.05	97.19	0.00	12050	55	224.73	0.00	1340.31	0.00	132550
6	50.58	0.00	181.98	0.00	14460	56	232.59	0.00	1279.62	0.00	134960
7	47.08	0.00	144.57	0.00	16870	57	214.06	0.00	1280.70	0.00	137370
8	30.64	0.00	182.47	0.00	19280	58	244.07	0.00	1420.85	0.00	139780
9	43.26	0.00	226.38	0.00	21690	59	256.32	0.00	1387.15	0.00	142190
10	54.63	0.00	266.56	0.00	24100	60	255.84	0.00	1435.64	0.00	144600
11	39.24	0.00	229.34	0.00	26510	61	253.22	0.00	1484.77	0.00	147010
12	62.77	0.00	287.93	0.00	28920	62	244.77	0.00	1460.85	0.00	149420
13	83.37	0.00	334.59	0.00	31330	63	275.16	0.00	1510.65	0.00	151830
14	37.71	0.00	323.75	0.00	33740	64	238.25	0.00	1506.03	0.00	154240
15	85.82	0.00	374.96	0.00	36150	65	291.62	0.00	1598.95	0.00	156650
16	79.28	0.00	406.70	0.00	38560	66	284.03	0.00	1613.87	0.00	159060
17	64.43	0.00	420.46	0.00	40970	67	304.09	0.00	1688.28	0.00	161470
18	94.44	0.00	518.19	0.00	43380	68	305.27	0.00	1643.49	0.00	163880
19	87.37	0.00	526.15	0.00	45790	69	302.74	0.00	1653.96	0.00	166290
20	90.35	0.00	449.95	0.00	48200	70	289.23	0.00	1695.45	0.00	168700
21	91.74	0.00	453.51	0.00	50610	71	278.63	0.00	1671.25	0.00	171110
22	114.59	0.00	503.94	0.00	53020	72	240.63	0.00	1693.60	0.00	173520
23	109.24	0.00	515.95	0.00	55430	73	313.67	0.00	1782.60	0.00	175930
24	87.69	0.00	574.08	0.00	57840	74	291.48	0.00	1796.27	0.00	178340
25	85.64	0.00	544.36	0.00	60250	75	305.02	0.00	1774.06	0.00	180750
26	94.83	0.00	608.02	0.00	62660	76	343.49	0.00	1874.32	0.00	183160
27	107.06	0.00	678.69	0.00	65070	77	317.44	0.00	1791.56	0.00	185570
28	94.91	0.00	632.28	0.00	67480	78	299.73	0.00	1816.46	0.00	187980
29	105.54	0.00	650.85	0.00	69890	79	322.27	0.00	1912.33	0.00	190390
30	135.86	0.00	747.07	0.00	72300	80	322.25	0.00	1906.08	0.00	192800
31	103.88	0.00	607.00	0.00	74710	81	305.26	0.00	1895.94	0.00	195210
32	136.23	0.00	778.83	0.00	77120	82	332.31	0.00	1932.23	0.00	197620
33	144.32	0.00	710.73	0.00	79530	83	333.46	0.00	2004.92	0.00	200030
34	162.65	0.00	839.06	0.00	81940	84	338.15	0.00	1989.05	0.00	202440
35	122.75	0.00	860.57	0.00	84350	85	353.79	0.00	2053.07	0.00	204850
36	117.85	0.00	811.34	0.00	86760	86	375.60	0.00	2081.86	0.00	207260
37	165.01	0.00	848.51	0.00	89170	87	341.61	0.00	2074.83	0.00	209670
38	197.77	0.00	873.63	0.00	91580	88	345.16	0.00	2044.23	0.00	212080
39	176.47	0.00	936.53	0.00	93990	89	358.02	0.00	2109.47	0.00	214490
40	173.56	0.00	991.87	0.00	96400	90	381.34	0.00	2152.92	0.00	216900
41	119.25	0.00	870.05	0.00	98810	91	374.45	0.00	2184.74	0.00	219310
42	179.95	0.00	996.62	0.00	101220	92	355.33	0.00	2184.31	0.00	221720
43	190.97	0.00	1017.49	0.00	103630	93	381.55	0.00	2223.12	0.00	224130
44	203.12	0.00	1061.24	0.00	106040	94	380.38	0.00	2258.34	0.00	226540
45	186.54	0.00	1076.07	0.00	108450	95	375.37	0.00	2246.93	0.00	228950
46	237.71	0.00	1167.16	0.00	110860	96	404.79	0.00	2280.45	0.00	231360
47	207.46	0.00	1189.84	0.00	113270	97	400.82	0.00	2307.26	0.00	233770
48	216.57	0.00	1177.11	0.00	115680	98	401.18	0.00	2310.08	0.00	236180
49	179.54	0.00	1196.76	0.00	118090	99	404.46	0.00	2359.05	0.00	238590
50	195.23	0.00	1167.21	0.00	120500	100	411.23	0.00	2380.17	0.00	241000

Table B.5: Kruskal Wallis tests with the sub-sampled datasets - Second round experiment

	Views		Gross Revenue				Views		Gross Revenue		
Id	χ^2	p-value	χ^2	p-value	sample size	Id	χ^2	p-value	χ^2	p-value	sample size
1	7.17	0.07	7.50	0.06	2420	51	158.47	0.00	75.33	0.00	123420
2	9.90	0.02	4.08	0.25	4840	52	172.02	0.00	96.60	0.00	125840
3	13.61	0.00	3.73	0.29	7260	53	138.89	0.00	59.86	0.00	128260
4	14.50	0.00	6.87	0.08	9680	54	145.13	0.00	59.04	0.00	130680
5	11.66	0.01	10.68	0.01	12100	55	168.68	0.00	82.65	0.00	133100
6	25.65	0.00	17.24	0.00	14520	56	170.88	0.00	80.53	0.00	135520
7	25.03	0.00	11.07	0.01	16940	57	146.29	0.00	71.80	0.00	137940
8	15.23	0.00	6.18	0.10	19360	58	196.64	0.00	94.13	0.00	140360
9	38.03	0.00	16.80	0.00	21780	59	143.70	0.00	62.34	0.00	142780
10	29.48	0.00	12.25	0.01	24200	60	160.44	0.00	80.16	0.00	145200
11	32.13	0.00	19.04	0.00	26620	61	172.13	0.00	70.57	0.00	147620
12	43.80	0.00	13.41	0.00	29040	62	203.61	0.00	88.55	0.00	150040
13	19.34	0.00	9.64	0.02	31460	63	190.19	0.00	84.31	0.00	152460
14	49.78	0.00	21.09	0.00	33880	64	168.42	0.00	68.10	0.00	154880
15	45.46	0.00	10.62	0.01	36300	65	199.86	0.00	77.13	0.00	157300
16	31.47	0.00	16.59	0.00	38720	66	192.17	0.00	93.86	0.00	159720
17	55.61	0.00	20.49	0.00	41140	67	217.58	0.00	95.90	0.00	162140
18	46.66	0.00	16.89	0.00	43560	68	203.38	0.00	84.40	0.00	164560
19	61.43	0.00	24.54	0.00	45980	69	189.13	0.00	83.47	0.00	166980
20	51.77	0.00	23.20	0.00	48400	70	169.79	0.00	80.95	0.00	169400
21	59.63	0.00	21.77	0.00	50820	71	208.43	0.00	104.34	0.00	171820
22	50.52	0.00	23.22	0.00	53240	72	208.94	0.00	93.78	0.00	174240
23	72.54	0.00	38.57	0.00	55660	73	222.98	0.00	107.28	0.00	176660
24	87.76	0.00	40.27	0.00	58080	74	234.23	0.00	96.72	0.00	179080
25	41.66	0.00	22.48	0.00	60500	75	231.43	0.00	107.60	0.00	181500
26	74.41	0.00	28.15	0.00	62920	76	205.96	0.00	81.68	0.00	183920
27	43.73	0.00	22.61	0.00	65340	77	194.53	0.00	87.82	0.00	186340
28	80.54	0.00	28.21	0.00	67760	78	208.95	0.00	84.71	0.00	188760
29	95.12	0.00	51.98	0.00	70180	79	234.43	0.00	94.95	0.00	191180
30	104.15	0.00	40.32	0.00	72600	80	236.12	0.00	102.21	0.00	193600
31	106.91	0.00	61.29	0.00	75020	81	230.93	0.00	95.35	0.00	196020
32	80.28	0.00	41.10	0.00	77440	82	224.18	0.00	96.63	0.00	198440
33	81.70	0.00	42.10	0.00	79860	83	238.28	0.00	93.96	0.00	200860
34	70.73	0.00	28.19	0.00	82280	84	249.63	0.00	106.99	0.00	203280
35	91.92	0.00	46.29	0.00	84700	85	239.48	0.00	105.61	0.00	205700
36	109.54	0.00	50.01	0.00	87120	86	227.71	0.00	101.11	0.00	208120
37	131.04	0.00	48.44	0.00	89540	87	264.08	0.00	116.04	0.00	210540
38	113.53	0.00	30.18	0.00	91960	88	268.40	0.00	118.96	0.00	212960
39	106.40	0.00	52.96	0.00	94380	89	245.56	0.00	106.17	0.00	215380
40	94.31	0.00	38.99	0.00	96800	90	235.94	0.00	101.99	0.00	217800
41	129.37	0.00	67.41	0.00	99220	91	256.21	0.00	111.49	0.00	220220
42	90.54	0.00	32.58	0.00	101640	92	250.41	0.00	108.99	0.00	222640
43	138.41	0.00	55.58	0.00	104060	93	282.25	0.00	127.52	0.00	225060
44	115.39	0.00	48.91	0.00	106480	94	271.14	0.00	117.16	0.00	227480
45	119.70	0.00	54.58	0.00	108900	95	276.68	0.00	117.18	0.00	229900
46	103.45	0.00	41.70	0.00	111320	96	273.06	0.00	121.59	0.00	232320
47	151.83	0.00	67.74	0.00	113740	97	279.50	0.00	121.42	0.00	234740
48	123.75	0.00	42.45	0.00	116160	98	277.22	0.00	118.71	0.00	237160
49	139.17	0.00	72.48	0.00	118580	99	277.24	0.00	120.47	0.00	239580
50	139.43	0.00	63.70	0.00	121000	100	284.65	0.00	123.16	0.00	242000