13

Understanding Media Markets in the Digital Age
Economics and Methodology

Brett Danaher, Samita Dhanasobhon,
Michael D. Smith, and Rahul Telang

13.1 Introduction

Digital distribution channels have created opportunities that have transformed the delivery of information, opening new ways for firms to add value to media and entertainment products. However, these new opportunities can create tension for firms struggling to adapt their business models to new markets and new competitors. The availability of pirated digital content only exacerbates this conflict, making it even harder for firms to develop viable digital business models. Piracy also raises issues for governments seeking to adapt established copyright practices to the unique realities of digital markets. Our intent in this chapter is to provide a tutorial for applying modern empirical methodologies to the abundance of natural experiments brought about by discrete changes in the media distribution market, thereby helping firms and governments adapt their practices based on data and empirical evidence as opposed to dogma and conventional wisdom.

Our position in this chapter is that empirical research using modern methods for causal identification are called for in order to determine the optimal

Brett Danaher is assistant professor of economics at Wellesley College. Samita Dhanasobhon is a PhD student at the H. John Heinz III School of Public Policy and Management at Carnegie Mellon University. Michael D. Smith is professor of information technology and marketing at the H. John Heinz III School of Public Policy and Management at Carnegie Mellon University. Rahul Telang is professor of information systems and PhD program chair at the H. John Heinz III School of Public Policy and Management at Carnegie Mellon University.

The authors thank the editors of this volume as well as Heidi Williams for helpful comments on this research. This research was conducted as part of Carnegie Mellon University's Initiative for Digital Entertainment Analytics (IDEA). Danaher acknowledges financial support from an Economics of Digitization Research Grant from the National Bureau of Economic Research. For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see http://www.nber.org/chapters/c12999.ack.

copyright and business policies in the digital era. Having written several papers on these subjects, we hope to provide a roadmap for future research to apply econometric methods for causal inference to answer the many managerial and policy questions raised by digital markets. The research we discuss in this chapter addresses these questions by exploring factors that influence demand for media content across various distribution channels and how consumers respond to changes in these channels resulting from firm strategy or government action.[1] Our research to date has shown the following: First, that a graduated-response antipiracy law in France causally increased digital music sales by 22–25 percent following widespread awareness of the law. Second, that the shutdown of the popular file-sharing cyberlocker Megaupload.com causally increased revenues from digital movies by 6–10 percent. Third, that the removal of NBC's video content from the iTunes store caused piracy levels of that content to increase by 11 percent but had no impact on DVD sales of the same content, implying that digital distribution of media may mitigate piracy without necessarily cannibalizing physical channel sales in the short run. Finally, new research in this chapter demonstrates that dis-tribution of television through online streaming (in this case, Hulu.com) can decrease piracy of that content by 15–20 percent. In short, our research seems to suggest that firms can compete against "free" pirated content by either making legitimate digital content easier to consume, or by making pirated content harder to consume. This implies that both firm strategy and government intervention may play a role in managing the disruption caused by digitization.

The remainder of our chapter proceeds as follows: In section 13.2 we summarize three of our prior studies pertaining to digital media, with a particular focus on the methodologies employed and other questions those methodologies might be used to answer. In section 13.3 we present new research on the impact of legal online streaming on demand for piracy. Finally, in section 13.4 we discuss the results presented in the chapter and set the agenda for future research.

13.2 Three Categories of Natural Experiments

In order to better understand the impact that a government intervention or a new firm strategy has on outcome variables such as sales or piracy levels, one must have a means to isolate and identify the causal impact of the event on the outcome. For example, if a government were to pass a policy aimed at reducing piracy, simply examining piracy levels before and after implementation of the policy would be insufficient to identify the impact

1. This chapter focuses on examples from our own work in order to describe various empirical methods. We are not the only ones to use such methods to explore questions related to the digitization of the media industries. For a broader summary of research and findings related to these issues, see Danaher, Smith, and Telang (2014).

of that policy change as piracy levels may have risen or fallen at that time for reasons unrelated to the policy. In the words of a common adage in the social sciences, "correlation doesn't establish causation."

To establish causation in such an environment, economists and social scientists often use a difference-in-difference strategy. The basic idea of a diff-in-diff approach is to identify a "control" group of individuals, regions, or products that can aid in estimating the counterfactual of what would have happened to the "treated" group if the treatment had not happened. The difference between this counterfactual and what we observe indicates the actual effect of the treatment, assuming that the control group can accurately predict the counterfactual. Thus, the selection of the control group is of paramount importance. The "gold standard" of causal inference in such research is randomized controlled trials (RCTs), whereby a random set of individuals or products are treated with a shock and the others are not. Such trials may not be out of reach—in our experience, firms in the media industries have been willing to randomly select some products to "treat" with availability on a new channel, shorter release windows, or variation in prices. When selection is truly random, many of the usual concerns about endogeneity are less salient, as unobserved characteristics will be similar on average across the control and treatment groups. Such experiments can be of value to both firms and researchers. However, when RCTs are not available, sometimes a natural or quasi-experiment can be found in which the selection of subjects into the treatment group may not be random, but may be random with respect to the outcome variables of interest. In this section, we give three examples of natural experiments and methodologies that can be used to analyze the causal impact of a treatment using a difference-in-difference methodology, but where each case involves a different type of variation in the data and thus a different manner of applying the methodology.

13.2.1 Case 1: The Effect of a Graduated Response Antipiracy Law on Digital Music Sales[2]

In the spring of 2009, the French government passed an antipiracy law known as HADOPI, establishing the HADOPI administrative authority and giving it the power to monitor online copyright infringement and to act against pirates based on information submitted by rights holders. The HADOPI authority had a number of responsibilities, including promoting and educating consumers about legal sales channels, but the most widely known program under HADOPI was the strikes and penalty system. Under this system, individuals would receive a warning for their first and second observed instances of copyright infringement, and upon the third they could be taken to court and potentially penalized with monetary fines or suspension of their Internet access for up to one month. This law was controversial

2. See Danaher et al. (2014).

and received a great deal of publicity, causing consumers to be very aware of the new policy and potentially affecting their behavior by migrating potential file sharers to legal purchasing channels. To analyze the impact of the HADOPI law on French consumers' digital music purchases, we obtained a panel of weekly iTunes digital music sales data from the four largest music labels for nearly three years surrounding the passage of HADOPI.

In this instance, the policy shock—the passage of HADOPI—was limited to one geographic region (France), and there was little reason to think it would have direct impact outside the boundaries of that country. Most other European countries had not experienced any relevant policy shocks at this time, and so our goal was to find a set of control countries whose sales trends over time closely matched France's prior to HADOPI, expecting that such a control group should have continued to trend similarly to France if not for the policy shock. We considered several control groups that in theory might have such a trend, examining only the pre-HADOPI sales trends to find the group that most closely matched France's trend.[3] The group of countries that best matched France's sales trends in the preperiod of our data was Spain, Germany, Italy, Belgium, and the United Kingdom. Notably these were also the five countries, other than France, with highest digital sales levels among European Union (EU) countries.

Before running a diff-in-diff model, another challenge that arose was selection of the "treatment date" that we would use in our model. Sometimes this is clear—if a government were to one day simply block all access to pirated material, that day would be the most obvious treatment date for analysis. However, the HADOPI bill was debated for over six months in the French government, even being passed by one government body only to be rejected and then subsequently accepted by another. With such confusion as to whether the law was in effect or not, we chose to consider the peak level of awareness of HADOPI as the effective treatment date. Google Trends data is a useful tool for measuring awareness of a law or policy, as it measures the number of searches over time for a given search term (as well as the number of articles containing the search term) for a given geographic area. Thus we used Google Trends to augment our data set and determine the effective treatment date of HADOPI.

The following Figure 13.1, reproduced from our paper, shows the results from an ordinary least squares (OLS) model predicting the natural log of iTunes song sales for France and the control group plotted against the Google Trends index of searches in France for the term "HADOPI."

3. In this stage, examining only the pre-HADOPI period is important. If one were to examine the entire period, one might be guilty of a form of "data mining," searching for a control group against which France would appear to increase or decrease after HADOPI. By only examining the preperiod to find the best-fitting control group, one remains agnostic as to the effect of the treatment and thus the diff-in-diff test that follows is valid.

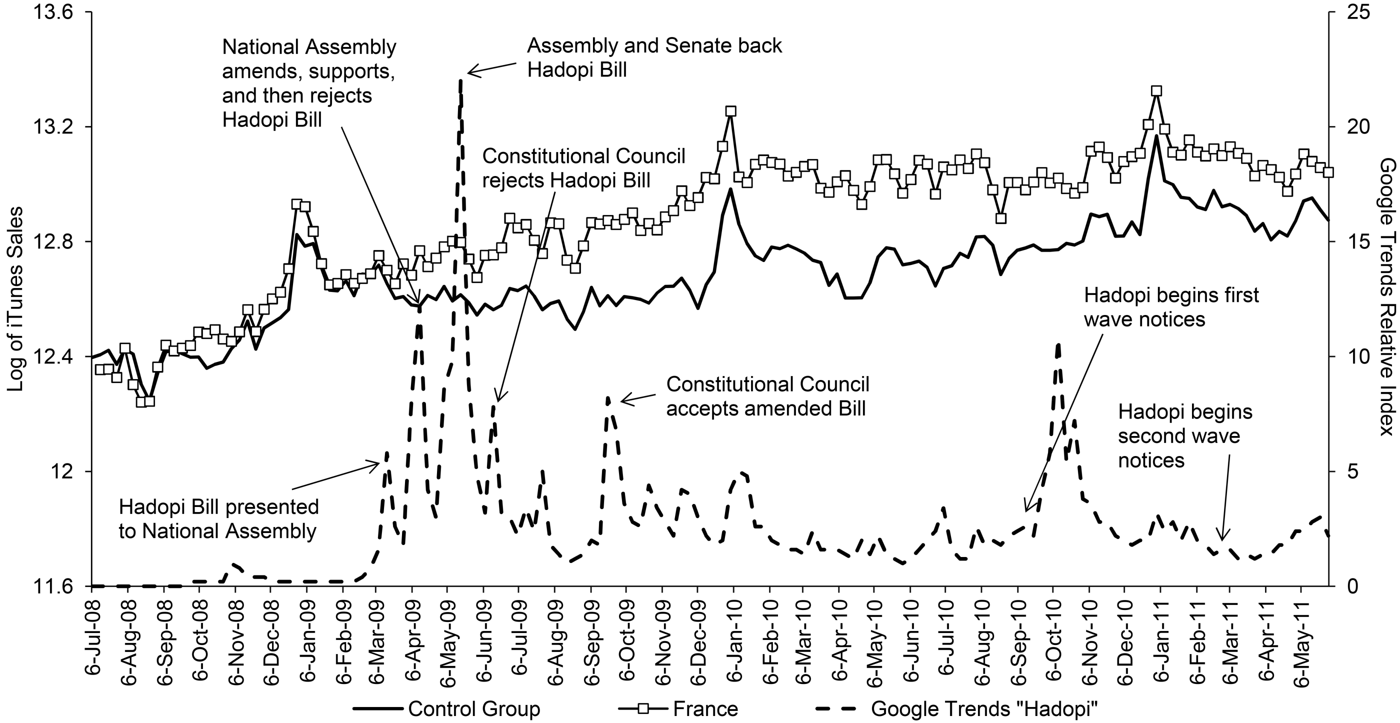

Fig. 13.1 **iTunes song unit sales trends, France versus control group**

Two important facts are clear in this picture. First, weekly sales trends of the control group match closely to sales trends in France prior to widespread awareness of HADOPI (moreover, a formal statistical test of joint differences between the control and treatment groups before treatment cannot reject that the two trends are the same during this period at a 95 percent confidence level). Second, increased awareness of the HADOPI law (proxied by Google search intensity) coincides with the persistent rise in the French sales trend above the control group.[4] Thus, these results suggest that awareness of the HADOPI law in France had a positive causal impact on iTunes sales in France, and that laws like this may migrate consumers from illegal file sharing to legal digital channels.

To provide further evidence that the effect we found was indeed causal, we added another level of difference to the model showing that the diff-in-diff increase in French sales was larger for more heavily pirated genres of music (and thus genres that should be more significantly impacted by the law) and smaller for less pirated genres. The logic here is that more heavily pirated genres should have a larger number of customers "treated" by the antipiracy intervention than less heavily pirated genres do.

The more general point about this paper is that when a government passes a policy or a firm implements a strategy in one region and when other regions could be expected to be unaffected by that change, a diff-in-diff strategy can provide useful evidence as to the policy's impact when a suitable control group can be found. This is not always easy. The iTunes store had been open in each of these countries for similar periods of time and so development of the market was reasonably stable across these countries. However, we found it difficult to study the impact of HADOPI on users of legal music streaming services like Deezer or Spotify, as these services were at very different levels of development across countries, and thus we could find no group of countries whose sales/subscription trends were following a pattern similar to France's. Despite this limitation, we believe that policy variation across countries (coupled with additional differences across attributes like genre) will be a powerful tool to analyze the impact of other government interventions like the Digital Economy Act in the United Kingdom and the Copyright Amendment Act of 2011 in New Zealand, as well as industry-led interventions like the Copyright Alert System put in place by US Internet Service Providers.

4. A point worth making about studies such as this is that the traditional standard error clustering approach (Bertrand, Duflo, and Mullainathan 2004) does not generate correct standard errors for the treated group in the posttreatment period, partly due to the low number of countries in the study, but also due to the fact that there is only one treated group. Our paper outlines a manner in which robust standard errors can be calculated in such a situation through permutational inference.

13.2.2 Case 2: The Effect of the Megaupload Shutdown on Digital Movie Sales[5]

In January 2012, the US Department of Justice secured an indictment against the popular cyberlocker Megaupload.com, allowing them to raid Megaupload's offices and shut down Megaupload's Internet presence. Prior to this, Megaupload was an online cloud storage service and the thirteenth most visited site on the Internet according to Alexa.com. However, according to the injunction, the vast majority of the content stored on Megaupload was copyright infringing and Megaupload's policies (such as not requiring passwords for storage accounts or providing incentives to upload popular content) encouraged rampant file sharing. The shutdown was controversial on many fronts, and opponents of the shutdown claimed that in spite of all of the costs of this government intervention, it would have little impact on consumer behavior as the content that had been available on Megaupload was available through other piracy channels (a conjecture aligned with empirical evidence presented by Lauinger et al. [2013]).

From an empirical perspective, what was notable about the shutdown was that it occurred all over the world on the same date and thus, unlike in our HADOPI study, there was no geographic region that could be considered a "control" area for estimating how sales would have changed in the absence of the shutdown. This challenge also arises with other policies or strategies that are taken worldwide all at once, or when there is a shock to a country but the only appropriate variation to study is within that country. In situations like this, no clear control group exists and so the simplest form of difference-in-difference may not be adequate to estimate the causal impact of the shock.

Fortunately, another way of implementing a diff-in-diff approach is to model the first difference as post- versus pretreatment but to use a more continuous variable as the second difference, where the continuous variable is a measure of how intensely each individual, region, or unit in the data was treated. In the Megaupload example, even though Megaupload was shut down in every country on the same date, each country had different preshutdown usage levels of Megaupload. To measure this variation, we gathered data on the number of unique visitors to Megaupload.com by country for the month prior to the shutdown, as well as data on the number of Internet users in each country at the end of the same month. Dividing the former by the latter, we imputed each country's Megaupload Penetration Ratio (MPR), or the percent of Internet account holders who visited Megaupload at least once in the month prior to the shutdown. With respect to the shutdown, the MPR can be seen as a measure of treatment intensity, as countries with higher MPR received a stronger "shock" from the shutdown

5. See Danaher and Smith (2014).

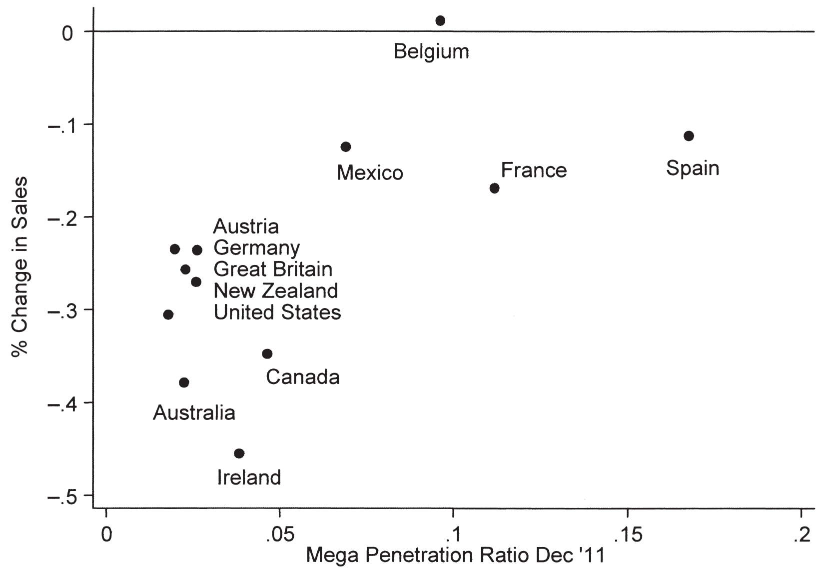

Fig. 13.2 Post-shutdown change in digital movie sales versus MPR (three weeks before and after shutdown)

and thus, if the shutdown actually boosted media sales, the post-shutdown sales growth should be larger in high MPR countries relative to low MPR countries.

Combining the MPR data with weekly digital movie sales data from two of the major motion picture studios, we showed that prior to the shutdown, the sales trends of high MPR countries were relatively similar to the sales trends of low MPR countries.[6] But immediately after the shutdown, high MPR countries experienced larger growth (or smaller declines from December to January sales levels) than low MPR countries do. Figure 13.2 presents a scatterplot that demonstrates this relationship, but in the paper we display results from OLS regression models that more precisely show the sales trends and more strongly support our inference that the shutdown of Megaupload caused an increase in digital movie sales.

One thing that stands out about this scatterplot is the positive relationship between increased MPR (x-axis) and increased relative sales change between December and January (y-axis). This positive relationship is the basis for the rest of the statistical evidence we provide in the paper that the shutdown of Megaupload caused an increase in digital movie sales. Another key takeaway is the importance of the diff-in-diff methodology here: sales in

6. The exception is during the Christmas holiday. In the paper we discuss how we deal with this anomaly in the preshutdown period.

nearly all of the countries were actually decreasing after the shutdown, but this is due to a seasonal decline from Christmas highs that happen every year in January. Simply examining average sales before and after the shutdown would show a decrease following the shutdown, but our diff-in-diff evidence indicates that the natural seasonal declines were mitigated by the closure of Megaupload, thereby causing revenues to be higher than they would have been if not for the closure.

It is worth nothing that in studies like this with a small number of clusters or "experiments" (countries), one might worry that preexisting trends could drive the results if high MPR countries were already growing faster than low MPR countries. In our paper we provide evidence from the preperiod indicating that this does not appear to be a driving factor. However, a better solution in situations like this is to add in country-specific trends to the diff-in-diff regression. Essentially this means modeling each country's specific week-to-week time trend based on some functional form (linear, quadratic, etc.), adding these terms into the regression, and asking if post-shutdown deviations from these modeled trends are larger in high MPR (high treatment intensity) countries. In the paper, we also showed that the addition of these trends actually increased the magnitude of our coefficient of interest and did not impact its sign or significance.

As an additional test of causal inference, we tested whether the relationship between MPR and sales changes was unique to 2011–2012 (when Megaupload was shutdown) or whether this same sales change pattern was common during this time of year. Indeed, in event studies such as these, a placebo test of a similar time period at some point (or in some location) where there was no treatment can help to verify causal inference. Accordingly, we showed that there was no statistically significant relationship between the December 2011 MPR and the percent change in digital movie sales after January 19, 2013.

Finally, from a policy perspective, one might ask how a model like this, one that uses variation in treatment intensity across regions, can be interpreted and explained to someone without training in econometrics. Essentially, what the model does is to model the linear relationship (or any functional form one considers appropriate) between pre-shutdown MPR and post-shutdown changes in sales. This relationship can then be extrapolated to estimate what would have been the post-shutdown sales change in a country with zero Megaupload usage, which is akin to asking what would have happened to sales in a country unaffected by the shutdown. In this manner a control "counterfactual" is estimated, allowing one to then estimate how much lower sales would have been in each country if not for the shutdown. An analogy could be made to a form of medical trial—the experiment is like giving one group of sick patients a pill that is 20 percent medicine and 80 percent sugar (placebo), giving another group a pill in a 40/60 percent ratio, and still another group an 80/20 percent pill, and then asking whether the groups given a higher concentration of medicine began to recover faster after the treatment than the groups given lower concentrations.

We suggest that the type of event study we conducted with Megaupload might also be useful for examining the effects of shocks when there is no clear control group. For example, the shutdown of Limewire in 2010 was similar to Megaupload, and its effect on sales of recorded music should be of interest to policymakers. Or, in 2009, Youtube.com chose to stop allowing individuals in the United Kingdom access to all premium music videos on their site due to a breakdown in negotiations with the British Performing Right Society. If there existed some geographic variance across the United Kingdom in pre-blackout usage of YouTube for music video watching, then this shock could be used to determine the effect of streaming music content (on YouTube) on sales or piracy of that content—a question that is currently of great interest to many parties involved in the music industry.

13.2.3 Case 3: The Effect of Digital Distribution of Television on Piracy and DVD Sales[7]

Considerable debate exists within the media industries around the use of new digital distribution channels such as paid download stores like iTunes and subscription streaming services like Spotify or Hulu. Proponents argue that such channels will more readily compete with illegal file sharing by offering consumers a more convenient legal means of acquiring content that includes a revenue stream to rights holders. Critics worry that such channels—often delivering lower profit margins—will cannibalize pre-existing channels with higher profit margins. With each potential channel the answers to these questions may be different, and yet they remain critical to determining the profitability of such channels or, in some cases, the size and direction of royalties that should be paid for the delivery of content.[8] But often these new channels are opened or closed with little evidence as to their effects on other channels.

Fortunately these questions can sometimes be answered, not using variance at the geographic level as above, but rather using variance at the product or firm level. Whether or not certain products are offered on these new channels is often based not on the piracy or sales levels of those products, but on contractual negotiations between rights holders and delivery channels.

For example, in early 2007 around 40 percent of all video content on the iTunes store was provided by NBCUniversal. Due to contract disputes related to iTunes pricing policies, NBC chose not to renew their contract

7. See Danaher et al. (2010).

8. For example, if users listening to a subscription music streaming service buy more music from existing channels, then perhaps royalties are unnecessary. But if these users buy less music, substituting streaming for purchasing, then the rate of sales displacement resulting from the service might be one determinant of the size of royalties that the streaming service should pay to rights holders.

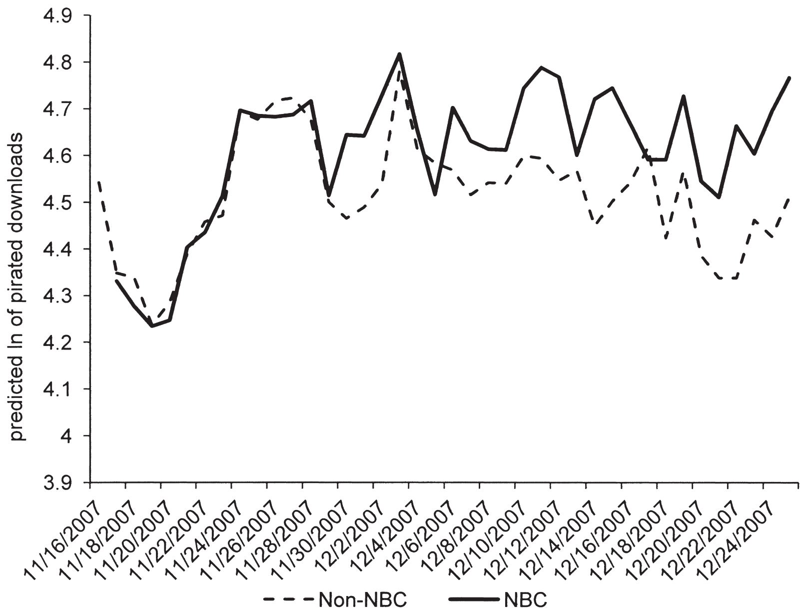

Fig. 13.3 NBC versus non-NBC piracy surrounding December 1, 2007

with iTunes and on November 30, 2007, they removed all of their television content from the iTunes store. However, similar networks (Fox, CBS, and ABC) continued to offer their content, providing a potential control group for NBC content. We used this product-level variation[9] and the NBC shock to determine the impact that selling television content on iTunes has on both piracy levels of that content and on physical DVD box set sales. Similarly, we used the return of NBC content to iTunes the following year to verify and provide additional insights into our results.

While the full results can be found in our paper, figure 13.3 highlights an example of the results from a diff-in-diff model comparing piracy of NBC content to a control group of ABC, CBS, and Fox content.

Similar to the results in our HADOPI paper, we show that the average pirated downloads of NBC episodes trended similarly to the average of control group episodes prior to iTunes removal,[10] but that immediately following the removal of NBC content from iTunes, piracy of those episodes spiked above the control group and remained above the control group during the

9. Technically this variation was at the network level, not the product level. But in the paper we argue that each television series was a unique experiment and treat standard errors accordingly.

10. An appropriate means of testing this is to ask whether a Wald test of joint significance for the difference between NBC and non-NBC content for all dates prior to the shock can be rejected at a specified significance level.

period covered by our data. Thus, we demonstrate that removing content from iTunes caused an increase in piracy, and by extension, that selling digitally on iTunes mitigates piracy. In the same paper, and using the same methodology, we showed that removal of NBC content from iTunes did not cause any increase in DVD sales of that content on iTunes, representing the reverse of the digital distribution question.

We believe that this approach has broad application to questions in the media industries in the age of digitization.

The negotiations between rights holders and content delivery platforms may create a plethora of natural or quasi-experiments where some rights holders come to terms with the platform (or do not come to terms) for reasons that can be shown to be unrelated to the dependent variables of interest. For example, on music streaming services, one label may choose to initiate or discontinue availability of its artists' albums while other labels make no changes to the status quo, and this might allow researchers to study the impact of music streaming on piracy, paid downloads, or CD sales. Our NBC paper provides a straightforward example of how to use such product-level variation to tease out the impacts of such strategies.

The focus of our descriptions of these three papers has been on the generalizability of these methodologies for a vast array of questions and experiments in the media industries following digitization. Specifically, our review establishes a set of methodologies and provides examples of how to impute causal impact across a variety of regularly occurring natural experiments—discrete changes at a country level (e.g., France and HADOPI), at a site level (e.g., Megaupload), or at a product or firm level (e.g., NBC and iTunes)—on variables of interest. Given the large number of these sorts of "natural experiments" driven by changes in how firms and governments respond to digital markets for entertainment, these methodologies could find wide application, and could help firms and governments understand the drivers of consumer behavior and the impact of such changes.

To demonstrate this, in the final section of this chapter we provide a proof-of-concept that these methodologies are generalizable to other settings by adapting the strategy from our NBC paper to study the effect of streaming television content on Hulu.com (a popular streaming site) on piracy of that content. Unlike the prior three examples where we provided high-level analysis, we now present precise details on data and methodology.

13.3 The Effect of Television Streaming on Piracy

Copyright holders have approached new digital distribution channels with a great deal of caution, despite the prevailing view that the vast majority of future sales inevitably will come through digital distribution, and the prevailing view that smart management should conduct experiments in advance of that arrival to understand the impact of these channels. Their concern about embracing new digital distribution channels seems to be driven by three main factors. First, digital distribution channels may substitute for sales in (more profitable) physical distribution channels. For example, Jeff Zucker, CEO of NBCUniversal, has been quoted as saying that the number one challenge for the motion picture industry in approaching digital channels is to avoid "trading analog dollars for digital pennies." Second, the use of digital distribution channels may accelerate the reduction in revenue from downstream channels, reducing the future profitability of present downstream channel partners. For example, it has been widely reported that Walmart forcefully protested Disney's distribution of its movies through iTunes by returning boxes and boxes of DVDs to Disney and by threatening to significantly reduce their future stock of Disney content. Finally, rights holders may be concerned that digital distribution channels are not commercially viable given the availability of "free" pirated content online. The concern here is that firms will have to significantly lower their prices today to compete with free pirated content and that this may reduce consumers' willingness to pay in the future. In short, competing with free pirated content today could have long-term impacts on the overall profitability of the industry in the future.

One managerial decision where these arguments have come into play is the decision of whether to allow television content to be shown on Internet websites for streaming video. Streaming video channels could be seen as low-margin competitors to the higher margin established broadcast of physical sales channels. On the other hand, allowing consumers to view television content through streaming channels may increase interest in the show and may decrease demand for digital piracy of this content. A legitimate streaming channel may also give copyright holders a great deal of flexibility in terms of assembling content and numerous opportunities to differentiate this content from physical DVDs, opening up new and untapped consumer markets and advertising revenues without significantly impacting demand in existing physical channels. In this more optimistic view, the firm who first figures out a viable streaming approach could improve its competitive position relative to its rivals, generating a strong incentive to experiment with these sorts of channels. Such a firm may also take a leadership position in creating platforms and infrastructure for digital distribution and streaming, thereby giving it a powerful position in the market.

Given these factors, it is notable that television and movie studios have begun to explore content distribution through many new digital distribution channels in recent years. These changes in distribution policies create a unique series of natural experiments in which to analyze the impact of free digital distribution on demand through physical channels and on demand for pirated content.

In our analysis below we analyze the impact of free streaming video websites on demand for digital piracy, and we also suggest that a similar approach could be used in the future to analyze the effect of streaming on physical sales or broadcast television. To analyze this question, we use a quasi-experiment that occurred on July 6, 2009, when ABC started streaming their television content on Hulu.com. Hulu.com is an advertising-supported Internet portal for streaming video. Interestingly, television networks themselves took leadership in creating this platform, and it was launched to the public in March 2008 as a joint venture between Fox and NBC. In April 2009, ABC reached an agreement to take a partial ownership position in Hulu.com and add its content to the site. This timing is important—Hulu had already existed for a year with content from two major networks, such that when ABC added their content to Hulu, the site already had a large existing user base and public awareness. As such, the addition of ABC represents a discrete shock to available content on a major delivery platform. The data suggests that this shock was exogenous with respect to piracy trends, as the timing was based on a series of contractual negotiations versus expectations of future piracy or sales.

In that sense, this experiment looks much like the one in our paper on NBC and iTunes in that when ABC added its content to Hulu.com on July 6, 2009, there were no shocks to content on other networks (NBC, CBS, CW, and Fox). Thus, television series on these four networks may serve as a control group for the treated ABC content, allowing us to identify the causal effect of Hulu.com streaming availability on levels of piracy. This differs from our prior paper on NBC in that we are studying a digital streaming service rather than a download service and we are studying the addition of content to a distribution channel rather than its removal from one.

Background and theory: Hulu.com was created as an attempt to give consumers a convenient, readily available platform on which to watch television content online on their own time. Unlike peer-to-peer file-sharing piracy, Hulu is a streaming service and requires no download time before one can watch episodes of a show. However, also unlike piracy, Hulu is supported by short, fifteen- to thirty-second advertisements inserted into the programs. And so despite the convenience and reliability of Hulu, it is not clear whether consumers will consider this service to be an attractive alternative to piracy.

During the timeframe covered by our data set, Hulu only offered the most recent five episodes of each television series, and all episodes and seasons before that were unavailable on Hulu.[11] Despite the fact that pirated copies of a television episode are often available through torrent sites the day after the episode airs, the owners of some series choose to delay availability of an episode on Hulu for several days after airing on television. This was not a factor in our study as the shock to availability occurred between seasons, so we study piracy of episodes of television that had aired at least a month prior to the beginning of our study. Nevertheless, it remains a question whether

11. Today, one can get access to all episodes of a number of series by paying to subscribe to Hulu Plus. However, Hulu Plus did not exist during the time period of our study.

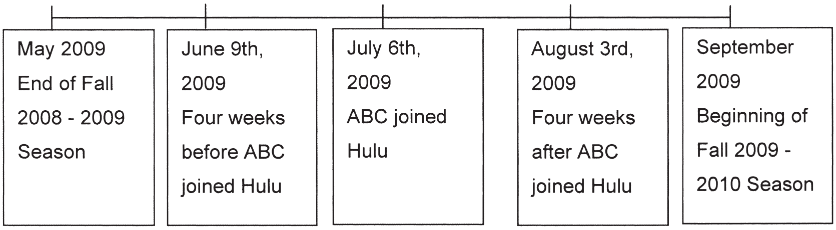

Fig. 13.4 Time line of events during period of study

consumers who would otherwise pirate will be attracted to the convenience (and legality) of Hulu enough to convert to consumption through legal streaming.

Finally, it is worth noting that television networks and their partners (like cable companies or downstream DVD sellers) may be worried that streaming would cannibalize DVD box set sales or over-the-air television viewing, where the profit margins are currently significantly higher than they are on streaming channels. In this study we will not analyze such potential cannibalization, but we believe one could undertake such analysis in the future with proper data on DVD sales, data on over-the-air audience viewing audience levels, and with a similar methodology to that employed here.

Data: To address the research question, we collected a panel of data on consumption of pirated television content through the BitTorrent tracker site Mininova.org. From these data, we analyzed all television series (excluding reality shows and live programming) that were available on the five major television networks (ABC, CBS, CW, Fox, and NBC) starting in the fall of 2008 and extending through the fall of 2009. This encompasses a total seventy-one television series. We describe these data in more detail below.

Figure 13.4 displays the time line of events in our study. It shows that ABC added its content to Hulu on July 6, 2009, a date after the end of the fall 2008 to spring 2009 television season and before the start of the fall 2009 to spring 2010 television season. As a result, we focus our analysis on episodes of television programs from the fall 2008 to spring 2009 season, and our analysis period covers the four weeks before and after ABC added its content to Hulu (with robustness checks for different window lengths). We also include only episodes that have at least ten downloads on each date to increase the signal-to-noise ratio of our tests.

Table 13.1 summarizes, by network, the seventy-one television series in our data and whether they were available on Hulu.com during the fall 2008 season. As noted in the table, of the seventy-one television series active in the fall 2008 to spring 2009 television season, twenty-seven of these series had their most recent five episodes available on Hulu.

In terms of what changed, note that prior to July 6, 2009, there were no ABC series available on Hulu, while after July 6, 2009, nine ABC television

Table 13.1 Hulu availability for each network's series, fall 2008–spring 2009 season

	Not on Hulu	On Hulu	Total
ABC	16	0	16
CBS	19	0	19
CW	6	2	8
FOX	6	8	14
NBC	6	8	14
Total	53	18	71

series became available on Hulu. These are the only changes in availability during this time frame—of the remaining sixty-two series, the forty-four that were not available on Hulu prior to July 6, 2009, remained unavailable after July 6, 2009, and the eighteen that were available on Hulu prior to July 6, remained available on Hulu after July 6. As such, from these television programs, we use the nine ABC television series that were made available on Hulu on July 6, 2009, as our treatment group and the remaining sixty-two series whose status on Hulu did not change as the control group.

Following Smith and Telang (2009) and Danaher et al. (2010), we use BitTorrent piracy measured by Mininova.org as a proxy for overall video piracy for the television content in our sample. We selected Mininova because it was the most popular BitTorrent tracker site listed by Alexa .com during our study period.[12] A further advantage of Mininova is that it provided the number of cumulative downloads for each tracker listed on its site, allowing us to calculate the number of daily downloads for each piece of content in our sample. The process for gathering these data and coding them are described in more detail in our NBC/iTunes paper.

To study the effect of the addition of ABC video content to Hulu, we focus our analysis on the four-week period before and after the July 6, 2009, launch date. This allows us to calculate the change in piracy for ABC content after its addition to Hulu.com, and to compare this change to the change in the control group. We focus our analysis on the four-week before and after period because we want to see the immediate impact of the policy and we want to exclude unrelated factors that might affect consumption over a longer time frame. We also test whether the change in piracy observed below is typical of other time frames by conducting the same analysis described here on the period one year prior to our study (the four-week period before and after July 6, 2008) as a further counterfactual reference point for how ABC piracy would have changed if it had not been added to Hulu. Importantly, we limit our piracy analysis to just the most recent five episodes of each series in our data, as these are the only episodes of any series (treatment or control)

12. See http://www.alexa.com/browse/general/?&CategoryID=1316737.

Table 13.2 **Daily number of downloads**

Pirated downloads		Mean	Std. dev.	Percent change
Treatment	Before 7/6/09	353.8	428.2	
	After 7/6/09	209.4	302.5	–40.80
Control	Before 7/6/09	388.4	558.5	
	After 7/6/09	301	437.7	–22.50

that were on Hulu.[13] Table 13.2 provides summary statistics of piracy data during the four-week period before and after July 6, 2009.

We use a balanced panel of episodes that were available both before and after ABC joined Hulu in these summary statistics and in our regression analysis. Table 13.2 reports the mean of the daily download numbers for the most recent five episodes of each series in both the control and treatment group. We found that the average number of daily downloads is consistent with the previous literature (Danaher et al. 2010), showing between 200 and 400 downloads per episode per day.

During the four-week period before and after the addition of ABC content to Hulu, the average number of daily pirated downloads for the last five episodes in the treatment group decreased by 40 percent, whereas the average number of daily pirated episodes for the control group decreased by 23 percent. We note that we would expect the number of downloads to decrease over time given that episode popularity will decline following an initial surge of interest immediately after broadcast. However, the relative sizes of these summary statistics suggest that there was a larger decrease in piracy for those series that were added to Hulu than there was for series where there was no change in their Hulu availability. We explore this result more formally in our regression analysis.

Results: Before comparing changes in the treatment and control groups after the introduction of ABC content to Hulu, we gather evidence as to whether piracy of the control group can be expected to trend similarly to piracy in the treated group if not for the shock. We use equation (1) to compare the time trend of piracy levels in the control and treatment groups prior to July 6, 2009. If the control group trends similarly to the treated group prior to the shock, then one might reasonably expect it to provide a good estimate of the counterfactual for the treated group after the shock.

13. It would certainly be interesting to consider the impact that having five episodes on Hulu would have on piracy of the entire series. However, sometimes individuals download a torrent containing all episodes from a season or series, and because of the nature of our observational data, we cannot determine whether the download of a season is because the downloading individual wanted just two to three of the most recent episodes, and downloaded the season torrent to get them, or actually wanted the entire season. Any analysis on piracy of episodes other than the five most recent would be subject to this data limitation.

$$(1) \qquad \ln Downloads_{it} = \beta_0 + \beta_1^t D_t + \beta_2^t D_t * ABCHulu_i + \mu_i + e_{it}.$$

In equation (1) above, $Downloads_{it}$ is the total number of pirated downloads of episode i on day t, D_t is a vector of date fixed effects for each day, $ABCHulu_i$ is an indicator variable equal to one if episode i is broadcast on ABC and was made available on Hulu on July 6, 2009 (and is equal to 0 for all episodes on other networks and untreated episodes on ABC), and μ is a vector of episode fixed effects. In this model, vector β_1 captures the day-to-day piracy trend for the control group, and β_2 represents how this differs for piracy of the treated group. Rather than displaying eight weeks worth of coefficients, we plot the predicted value from the resulting coefficients in figure 13.5 using $\beta_0 + \beta_1$ as the predicted log piracy of the control group and $\beta_0 + \beta_1 + \beta_2$ as the predicted log piracy of the treated group.

While figure 13.5 demonstrates that piracy trends of the treatment and control groups were not quite the same prior to the experiment, they were quite close. However, after treated ABC series were added to Hulu on July 6, 2009, there is an immediate break in piracy levels of the last five episodes of each of these series in the treated group that is much larger than any drop/change in piracy of the control group. Based on the timing of this relative drop and the lack of a similar drop before the experiment, we believe the

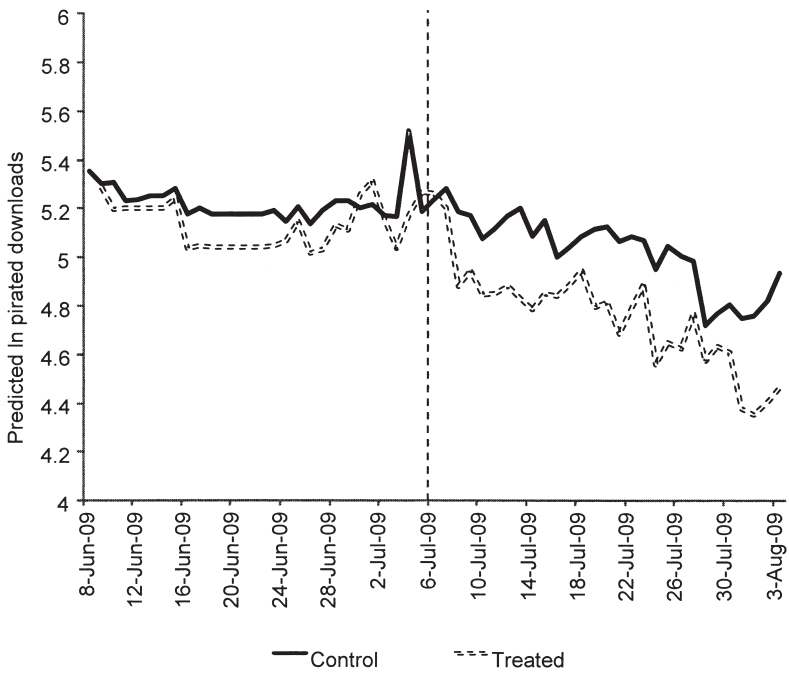

Fig. 13.5 Treated versus control group piracy surrounding July 6, 2009

most logical explanation is that people pirated ABC content less once it was added to Hulu.

In addition to this break in the treated group's piracy relative to levels in the control group, one also notes a break in the control group's piracy levels relative to historical norms. Because the other networks made no major policy changes on this date, this break might suggest a spillover effect: If new viewers of ABC content on Hulu discovered the other shows they like on Hulu, they may have stopped pirating those shows or they may have substituted from non-Hulu shows (which they previously pirated) toward newly discovered shows on Hulu. While we do not have a suitable identification strategy to formally test for these effects, we note that such a spillover effect result would be consistent with similar results in Danaher et al. (2010). They found that when NBC removed their television content from iTunes, in addition to an increase in demand for NBC piracy relative to the control group (ABC, CBS, FOX), there was also an increase in demand for piracy of the control group. Finally, we note that if there was a spillover effect in our present Hulu context, then our control group was partially impacted by the treatment and our reported results will underestimate the effect of adding content to Hulu on piracy of that content.

In order to obtain a statistical estimate of the size of the impact that the streaming channel had on piracy of ABC treated content, we adapt equation (1) as follows:

$$(2) \qquad \ln Downloads_{it} = \beta_0 + \beta_1 After_t + \beta_2 After_t * ABCHulu_i + \mu_i + e_{it}.$$

The variables in equation (2) are the same as in equation (1), except that here *After* is an indicator variable equal to one for all dates after and including July 6, 2009. Variable β_2 thus measures the average difference between treatment and control group in the period after ABC was added to Hulu, compared to any difference beforehand. Under the assumption that the treated group would have trended similarly to the control if not for the experiment, β_2 measures the effect that adding ABC content to Hulu had on piracy of that content.[14] Because there could exist correlation between downloads of different episodes of the same season or even series, we cluster our standard errors at the series level, treating each series in our data as a unique experiment.

Estimating equation (2) through OLS, β_2 is –0.19 (in the eight-week window specification), indicating that the postexperiment decrease in pirated downloads was 18 percent larger for treated ABC content than it was for

14. We ran a more flexible model with a full vector of date fixed effects that produced nearly identical estimates and standard errors for the coefficient of interest. But in this model the "after" variable (for the control group) is subsumed by these fixed effects and so we present the results from the less flexible specification in the table so that the reader may compare the change in the treatment group to the change in the control group.

Table 13.3 OLS of log-pirated downloads

	Eight-week window	Four-week window	Two-week window
After 7-6-2009	–0.194***	–0.072	–0.067
	(0.053)	(0.054)	(0.048)
After 7-6-2009 * ABC	–0.190	–0.169*	–0.164*
	(0.121)	(0.098)	(0.088)
Constant	5.214***	5.218***	5.232***
	(0.026)	(0.025)	(0.024)
Observations	14,132	7,121	3,886
No. of Series	71	71	71
R-squared	0.139	0.071	0.074

Note: Robust standard errors clustered at series level appear in parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

control content (see table 13.3). The p-value for this coefficient is 0.13, so we cannot reject the null hypothesis that changes in ABC piracy were the same as for the control group. This may be due to lack of power in the test: when we conservatively cluster standard errors at the series level, there are only nine treated clusters in the data. However, if we shorten the experimental window to either one or two weeks before and after the treatment (thereby reducing random variance from other unrelated factors), we find similar coefficients but with p-values less than 0.1, allowing us to reject the null hypothesis at a 10 percent significance level or lower.

As further evidence, we estimated equation (2) for the same dates in 2008 (using content from the fall 2007 season), expecting no diff-in-diff change for ABC content as there was no shock to content in this period. Indeed, β_2 for the 2008 period is estimated as –0.02 with a standard error of 0.04, indicating that the change in piracy of ABC content was economically and statistically insignificant relative to the change in piracy of the control group content in this placebo test.

While the significance levels are somewhat low due to small sample size, the magnitude of the estimate is fairly large. Thus our point estimates and our placebo test indicate a pattern in which the addition of ABC content to Hulu caused a nearly 20 percent drop in pirated downloads of the added content, and we interpret this result similarly to the results in our paper on NBC and iTunes. That is, delivering television content in more convenient, readily available channels can cause a substantial number of pirates to turn from illegal file-sharing channels to legal channels. Future research might explore the coding of the torrent data differently in an attempt to determine whether the addition of the most recent five episodes of a series to Hulu reduces pirated consumption of just those five episodes (our finding) or pirated consumption of the entire series.

13.4 Discussion

We began this chapter by pointing out a variety of questions that have arisen in the media industries as a result of the digitization of content and of the resulting weakening of intellectual property due to file sharing. The goal of this chapter was to point researchers to a number of topics that we believe to be interesting and of managerial or regulatory importance, and then to highlight the importance of using natural experiments that arise in the context of rapidly changing media markets as a way of addressing these and other related questions.

To this end, we have shown how several of our papers address these topics through the analysis of natural experiments and through exploiting different types of variance in the data. We have given suggestions of other government interventions or firm strategies that are not well understood and that could be studied with one of the methodologies from our prior work.

Finally, as proof of concept, we applied the difference-in-difference model from our paper on distribution through the iTunes channel to a completely different data set and event: the streaming of television content to consumers on Hulu.com. As file sharing continues to be a commonly chosen consumption channel and as firms continue to innovate through new platforms or strategies for delivering content, the ability to understand the interactions between these channels and the impact that government policies can have on digital markets will only increase in importance. We hope that this chapter serves as a basis for new research to paint a clearer, more complete picture of the complex interplay between media firms' strategies, government policy, and consumer behavior.

References

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-In-Differences Estimates? *Quarterly Journal of Economics* 119 (1): 249–75.

Danaher, Brett, Samita Dhanasobhon, Michael D. Smith, and Rahul Telang. 2010. "Converting Pirates without Cannibalizing Purchasers: The Impact of Digital Distribution on Physical Sales and Internet Piracy." *Marketing Science* 29 (6): 1138–51.

Danaher, Brett, and Michael D. Smith. 2014. "Gone in 60 Seconds: The Impact of the Megaupload Shutdown on Digital Movie Sales." *International Journal of Industrial Organization* 33:1–8. http://ssrn.com/abstract=2229349.

Danaher, Brett, Michael D. Smith, and Rahul Telang. 2014. "Piracy and Copyright Enforcement Mechanisms." In *Innovation Policy and the Economy*, vol. 14, edited by Josh Lerner and Scott Stern. Chicago: University of Chicago Press.

Danaher, Brett, Michael D. Smith, Rahul Telang, and Siwen Chen. 2014. "The Effect of Graduated Response Anti-Piracy Laws on Music Sales: Evidence from an Event Study in France." *Journal of Industrial Economics*. 62 (3): 541-553.

Lauinger, Tobias, Martin Szydlowski, Kaan Onarlioglu, Gilbert Wondracek, Engin Kirda, and Christopher Kruegel. 2013. "Clickonomics: Determining the Effect of Anti-Piracy Measures for One-Click Hosting." Working Paper, Northeastern University.

Smith, Michael, and Rahul Telang. 2009. "Competing with Free: The Impact of Movie Broadcasts on DVD Sales and Internet Privacy." *Management Information Systems Quarterly* 33 (2): 312–38.