THE PERFECT STORM:
Using Snowstorms to Analyze the Effect of Theatrical Attendance on the Demand for Subsequently Released DVDs

Patrick Choi    Peter Boatwright    Michael Smith
June 24, 2016

ABSTRACT

Movies are distributed through multiple, carefully segmented, channels. This paper investigates how consumption in a movie’s theatrical channel affects demand in the subsequent DVD retail channel. We exploit exogenous variation in events that affect theatrical attendance in a geographic market to estimate the causal impact of theater attendance on home entertainment demand. Specifically, we use the occurrence of major snowstorms surrounding a movie’s theatrical opening weekend as an exogenous shock to theatrical demand in a local market.

Using this instrumental variable in a control function framework, we find evidence that theatrical attendance causally impacts home entertainment demand: lower theatrical attendance in a geographical market that experiences an opening weekend snowstorm leads to lower DVD/Blu-ray sales in the movie’s subsequent home entertainment release window in that geographical market. Specifically, we estimate a 10 percent drop in theatrical attendance causes an approximate 2.7 percent decrease in the volume of DVDs/Blu-ray discs sold in the first eight weeks of the DVD release window. This result provides important managerial guidance in an industry undergoing significant changes in the how movies are marketed across theatrical and home entertainment channels.
1. Introduction

The marketing environment for motion picture content has changed significantly in recent years. While movies are almost always released first in theaters, and later in home entertainment formats such as DVD/Blu-ray discs, the importance of these home entertainment channels has increased significantly over time, both in terms of revenue and consumer interest. For example, theatrical revenue made up 55 percent of a typical movie’s revenue in 1980, but only 20 percent in 2007 with the remaining 80 percent coming from home entertainment releases (Epstein, 2012). In terms of consumer interest, a 2005 Ipsos survey found that only 22 percent of Americans surveyed would prefer to see a movie in a theater versus watching the same movie at home on DVD (Keating, 2012). More recently, theatrical attendance hit a two-decade low in 2014 (McClintock, 2014), the same year that ticket prices hit an all-time high (Linshi 2015). The increasing importance of the home entertainment window is also reflected in the changing marketing environment for home entertainment content, notably the reduced delay between average theatrical and DVD release dates, which declined from just under 6 months in 1998 to just under 4 months in 2013 (Ulin 2013) (see Figure 1 for a summary of a movie’s release timeline).

Given these changes, it is important for marketing managers to understand interactions between the theatrical and home entertainment channels. In this regard, although it is well known that a movie’s theatrical revenue is a strong predictor of its subsequent home entertainment revenue, there is no rigorous empirical evidence indicating whether increased theatrical attendance complements or substitutes for home entertainment demand. From a theoretical standpoint, theatrical attendance could have either effect: To the extent that consumers perceive the theatrical experience to be relatively undifferentiated from watching a DVD or Blu-ray disc at home, one would expect that the two channels would be substitutes — with increased consumption in one channel reducing demand in the other channel. However, if the channels are significantly differentiated, they could complement each other in a variety of ways, as hypothesized by Hennig-Thureau et al. (2007). To the best of our knowledge, there is no empirical research that estimates the magnitude of these complementary effects from field data. Therefore, whether the complementary
forces for theatrical consumptions on downstream DVD channels can outweigh the cannibalization remains an open question.

However, empirically testing whether theatrical viewership has a positive or negative impact on demand in subsequent distribution channels is challenging. Using observed theatrical admission and DVD/Blu-ray sales data to test the impact of theatrical attendance on DVD/Blu-ray demand at a movie level suffers from obvious endogeneity problems: unobserved movie popularity factors impact both theatrical demand and home entertainment demand in ways that available control variables do not capture. Because movies with superior popularity factors have higher demand in both theaters and home entertainment formats, analyses that do not account for these unobserved confounders would incorrectly attribute this correlation in demand to the effect of theatrical viewership on the demand for DVD/Blu-ray releases. To accurately test whether theatrical viewership has a causal impact on subsequent DVD/Blu-ray sales, we need an exogenous shock to theatrical viewership. Exogenous shocks introduce changes to theatrical viewership that are independent of all unobserved factors, and thus enable us to identify how changes in theatrical viewership affect subsequent DVD/Blu-ray sales.

In this paper we use major snowstorms surrounding a movie’s opening weekend as just such an exogenous shock. Major snowstorms impede travel and reduce theater attendance. The negative correlation between snowstorm occurrences and theatrical viewership, coupled with the random and unpredictable nature of snowstorm occurrences, produce plausibly exogenous variations in theatrical viewership across geographic markets for movies released in the winter. We then use this exogenous variation in theatrical attendance to determine how lower theatrical attendance in a particular geographical region impacts demand in the subsequent DVD/Blu-ray release window.

Our results show that theatrical demand causally increases DVD/Blu-ray demand. Specifically, a 10 percent decline in theatrical attendance causes a 2.7 percent decline in DVD/Blu-ray demand. This result suggests that the complementary forces outweigh the cannibalization in these two channels.
2. Related Literature

Our research is related to a variety of papers in the academic literature analyzing movie sales in the theatrical and home entertainment windows. For example, Lehmann and Weinberg (2000) specify a model that uses observed theatrical sales to predict video rentals in the home entertainment channel. Their paper specifies exponential curves for both theatrical sales from the theatrical channel and from the video rental channel. However, it is important to note that their paper focuses on predicting rental sales, not on establishing a causal relationship between theatrical attendance and video rentals. Thus, because their the paper does not account for unobserved confounders that affect demand in both distribution channels, it does not establish that a change in theatrical attendance would lead to a change in demand in subsequent home entertainment channels.

In a related study, Mukherjee and Kadiyali (2011) model the demand for DVD purchases and DVD rentals. Our paper differs from their study in that two channels modeled in Mukherjee and Kadiyali (2011) overlap, and thus consumers make simultaneous consumption decisions for the two channels, whereas the channels considered in this paper and Lehmann and Weinberg (2000) are separated temporally, allowing for sequential consumption decisions. Mukherjee and Kadiyali (2011) share a limitation similar to that in Lehmann and Weinberg (2000)—that unobserved demand shocks, such as unobserved movie popularity factors, confound their results. Neelameghan and Chintagunta (1999) model the box office performance of the US and international theatrical channels. They specify that viewership in each channel follows a Poisson distribution, and then link the mean parameters to control variables and movie characteristics in a hierarchical Bayesian specification. Again, unobserved movie popularity factors not fully explained by the control variables and observed movie characteristics would confound any conclusion on the substitution or complementarity nature of the channels. Finally, in an analysis of the advertising responsiveness in the US DVD market, Luan and Sudhir (2010) report that a 0.96% increase in DVD sales is associated with a 1% increase in the box office. Despite the comprehensive approach to handle the endogeneity issues in advertising spending, DVD release lag, and DVD retail price, their model was not designed to resolve the endogeneity problem in the box office for the determinant of DVD sales caused by omitted confounders.
Therefore, the positive association between box office and DVD sales reported in Luan and Sudhir (2010) does not establish that the two channels are complementary.

Our study is also related to two studies analyzing movie distribution in multiple sequential channels. Hennig-Thurau et al. (2007) suggest that a multiple-purchase effect, an information-cascading effect, and an uninformed-cascading effect can cause a potential complementarity between the theatrical channel and home entertainment channels. A multiple-purchase effect means consumers see a movie more than once, and their theatrical viewing stimulates the purchase in subsequent channels. An information-cascading effect means the success of the theatrical channel affects the performance of subsequent channels, through shared personal experience, such as word-of-mouth. An uninformed-cascading effect means the success of the theatrical channel affects the performance of subsequent channels through aggregate facts, such as released box office numbers. Calzada and Valletti (2012) constructed a game-theoretic model of movie distribution and consumption. An important implication of their model is that the optimal distribution strategy of movie studios depends on the substitutability among channels. If channels are strong substitutes for each other, the optimal distribution strategy should be sequential. On the other hand, if channels are weak substitutes, or complements, and consumers can buy from multiple channels, the optimal distribution strategy should be simultaneous release with reduced prices.

Our research extends both streams of the literature by first using an exogenous shock in theatrical viewership to establish a causal relationship between theatrical viewership and home entertainment demand, and second by providing empirical evidence to inform models, such as Calzada and Valletti’s, regarding the substitutability between these two important sequential release channels for movies.

3. **Mechanisms**

Hennig-Thurau et al. (2007) suggest three dominant mechanisms behind the finding that higher theater attendance causes higher DVD sales:

1. The multiple-purchase effect: a consumer’s in-theater consumption of a movie simulates his/her purchase of the DVD. Learning could cause this effect—information on the quality and taste
matching is revealed to a consumer when he/she watches the movie in the theater, and the revealed information reduces uncertainty. Later, when the consumer contemplates which movie to choose for DVD purchase for their own consumption or collection, he/she is more likely to purchase the DVDs of the movies with less uncertainty than those about which he/she has less information.

2. The information-cascading effect: in-theater consumption of movie increases the likelihood of a consumer spreading word-of-mouth; after watching a movie in the theater, a consumer may tell others in his/her local social circle about this movie and raise awareness for the movie in the geographic market. This higher level of awareness in turn leads to stronger sales in the DVD release window.

3. The uninformed-cascading effect: higher posted box office numbers from a more successful theatrical release creates higher awareness in the market, and in turn leads to higher demand for its DVD.

To investigate the relative plausibility of these three mechanisms in our setting, we conducted an online survey on the consumer theatrical and DVD purchase histories for movies (see Appendix A for the list of survey questions). Our survey was conducted through Amazon’s Mechanical Turk (n=223). We asked respondents to report the number of movies they had seen in theaters, and the number of DVDs they had purchased, in the last five years. We then inquired about the percentage of DVDs they had purchased after seeing the movie in theaters. In addition, we asked them to provide reasons why they buy the DVDs for movies they have already seen in theaters. These survey questions aim to test for the existence of a multiple-purchase effect. We also asked the respondents the percentage of DVDs they had purchased because of word-of-mouth from friends, and the percentage of DVDs they had purchased simply because the movie was a huge box office success. These two survey questions aim to investigate the existence of an information-cascading effect and an uninformed-cascading effect.

Of our 223 respondents, 70% had purchased DVDs in the last 5 years for movies they had seen in theaters. Eighty percent of these respondents stated that one of the key reasons they purchased DVDs after
seeing the movies in theaters is to re-watch it and 25% of these respondents stated that they purchased the DVD as a gift for friends and family (respondents were allowed to choose multiple reasons). Furthermore, excluding the respondents who purchased few DVDs (one or two DVDs in last five years), we find that 12% of all the purchased DVDs for respondents in our sample were for movies that consumers saw in theaters. This result is consistent with the existence of the multiple-purchase mechanism, because the survey shows consumers occasionally buy DVDs for movies they watched in theaters. On the other hand, 22% and 13% of all the DVD purchases were motivated by word-of-mouth from friends and by awareness generated by the movie’s box office success, respectively, in deciding to purchase a DVD. These results suggest that informed-cascading and uniformed-cascading effects may also drive the observed positive spillover from the theatrical channel to the DVD retailing channel.

An alternative explanation for our empirical result in our analysis below is that our finding of higher theatrical viewership leading to higher DVD sales is not driven by consumer behaviors, but rather by firms’ strategic actions. That is, movie studios and DVD retailers set their DVD pricing and advertising strategies based on the box office performance, and these strategic actions based on observed box office performance cause changes in DVD sales. However, this alternative explanation is unlikely to be valid in our setting. This paper uses market-level data to analyze the effect of theatrical viewership on DVD sales, and thus this alternative explanation would suggest studios and retailers set their DVD marketing-mix variables at a city or regional level as a reaction to the local box office performance. We reached out to two executives at the data-providing movie studios, and they stated that their studios do not set DVD marketing strategy at the local market level in response to theatrical popularity in that city.

4. Data

This paper uses DVD/Blu-ray sales and box office data from three major US movie studios. We use the movie’s box office gross revenue divided by the national average movie ticket price in the year of release as a proxy for theatrical attendance. The three participating movie studios provided data for movies
from different but overlapping release years: 2003–2012, 2006–2013, and 2011–2014. We exclude several 2014 movies that had incomplete DVD/Blu-ray sales data at the time of data delivery.

To maintain a relative homogeneity across titles, we focus on wide-release movies—movies that had more than 600 opening theaters in the United States, because platform releases (movies released in a small number of theaters initially) are systematically different than titles released using the (more common) wide release strategy. We also exclude foreign films that were released internationally several months to a year earlier than in United States, because these movies are fundamentally different than the US-produced movies, and also the higher availability of pirated copies from early international releases might affect the box office and DVD/Blu-ray sales.

The unit of analysis is the sales of a movie in a DMA. We have a total of 21,238 observations from 104 movies in 209 DMA markets. For each movie-market unit, the dependent variable is the total number of DVDs and Blu-ray discs sold through three big-box retailers (Walmart, Target, and BestBuy). Following the work of Eliashberg and Shugan (1997), Basuroy, Chatterjee, and Ravid (2003), and Liu (2006), we use a window of the first eight weeks for the sales of both theatrical and DVD/Blu-ray releases. The box office receipts of blockbuster-type movies decay exponentially over time (Ainslie, Drèze, & Zufryden, 2005), and receipts from the first eight weeks of theatrical release on average account for more than 95% of the box office revenue from the entire theatrical release window. We find that the volume of DVDs/Blu-ray discs sold over time follows a similar exponential decay pattern for the first three to four weeks and then stabilizes to a small stream of sales from the fourth week onward. Because the demand in both channels is heavily concentrated in the early weeks, analyses using the first eight weeks of sales are reasonable.

Table 1 presents the variable descriptions. Table 2 lists the summary statistics and pairwise correlations among the variables of interval scales. In the following section, we discuss each of the explanatory variables in detail.

**Explanatory variables at movie-market level:**

1. **Theatrical attendance:** We estimate attendance by dividing the total box office revenue from all theaters in the market for the movie in the first eight-week window by the national average movie
ticket price in the year of release. We include DMA fixed effects in our models to resolve the issue of variation in ticket prices across DMAs.

2. **Snowstorm instruments:** We use an opening-weekend-snowstorms instrument and a prior-week-snowstorms instrument. The opening-weekend-snowstorms indicator is set to one if any severe winter event occurred in the geographic market during the theatrical opening weekend; the prior-week-snowstorms indicator is set to one if any severe winter event occurred during the seven-day window before the day of the theatrical opening. A severe winter event is defined as a report of a Blizzard, Heavy Snow, Ice Storm, Winter Storm, or Winter Weather in the Storm Events Database from the National Oceanic and Atmospheric Administration’s National Climate Data Center. The records in the Storm Events Database are at the county level. Because a DMA can be comprised of multiple counties, we choose the county seat of the largest city in the DMA when we merge the county-level weather data with the DMA-level sales data. The severe-weather-event records are based on reports from various local sources such as the Park or Forest Service, trained spotters, and emergency managers. Because the severe winter events are based on trained personnel in the local area, these snowstorms are adjusted for snowfall in the local area. In other words, four inches of snowfall overnight may trigger a heavy snow event in a warmer-temperature city but may not trigger the same event in a colder-temperature city that is more accustomed to snow. One limitation of this database is that the availability of these reports varies over time as some sources went inactive and new sources were added. To alleviate the impact of the changing availability of sources, we filter out DMAs that have less than four years of severe winter events between 2003 and 2014. Our results are robust to alternative restrictive filtering schemes based on the restriction criteria of requiring five, six, or even more years of winter event reports.

**Explanatory variables at the movie level:**

3. **Movie characteristics:** We collected data on movie characteristics including production budget, advertising expense, number of opening screens in the United States, studio, genre, MPAA rating, the presence of star actors, and IMDB user-review rating. We obtained data on production budget,
number of opening screens in the United States, studio, genre, MPAA rating, IMDB user-review rating, and year and month of theatrical release from IMDB and Boxofficemojo websites. The indicator variable for star actors is set to 1 if any of the movie’s cast is in IMDB’s STARmeter Top 10 list the year of and the year immediately after theatrical release. IMDB’s STARmeter is designed to capture the level of public interest in an actor or actress based on the frequency with which his or her profile is viewed on the site. This variable is comparable to the Hollywood Reporter’s Star Power Index, which is used by other papers in the literature to control for the presence of star actors (Elberse & Eliashberg, 2003; Gopinath, Chintagunta, & Venkataraman, 2013).\(^1\) We use the presence on two consecutive years’ lists to determine whether an actor or actress is considered a major star, because lags may exist between the rise of a star and the year the new star appears on the IMDB list. Our advertising expense data was obtained from Kantar Media for each movie in our data. We use the year and the month of theatrical release, and whether the movie was released during Christmas school holidays (between December 23 and January 2) to control for the timing of movie releases. We also note that movie studios strategically choose the timing of theatrical openings based on revenue expectations. For example, movies with lower commercial expectations are more likely to be released in January than in other winter months. By including calendar month fixed effects in our model, we control for these release-timing strategic effects because the model effectively considers only variations across movies within the same calendar month. We also include year fixed effects to remove the confounding effects of economic cycles and other time trends. Lastly, to control for the magnitude of competition of a movie in a theater, we use the total production budgets of the movies released the same week as the focal movie. This variable is similar to the control of competition for “screen space” from new releases used in Elberse and Eliashberg (2003).

4. **DVD price at release:** We control for the price of the DVD at the time of its release because DVD price may be a factor in a consumer’s DVD purchase decision. Although the best control is the price

---

\(^1\) We used the IMDB STARmeter measure in our paper instead of the Hollywood Reporter Star Power Index because the most recent Hollywood Reporter star-power ranking was published in 2006, well before our study period.
of the DVD at DVD release in the local market, we do not have data at that level of granularity. Therefore, as a control, we use the median value of the daily prices over the first two weeks on Amazon.com. We divide the national DVD sales volume in the first week of DVD release by the national DVD unit sold in the first week of DVD release to estimate the national average price of the DVD at release.

5. Number of weeks between theatrical and DVD releases

We control for the length of delay of the release of the DVD after the release of the movie in theater, by computing the number of weeks in between the theatrical opening date and the DVD release date.

6. Characteristics of competing DVDs at the week of DVD release

To control for the magnitude of competition during the DVD release, we use the total production budgets of the new DVDs released the same week as the focal movie.

5. Model specification and Empirical Strategy

We postulate a flexible system of the relationship between theatrical attendance and subsequent DVD/Blu-ray sales volume for movie $m$ in DMA $d$:

$$
\bar{Y}_{md} = M_{yr(m),d}^{DVD} Y^*_{md} \\
Y^*_{md} = \tilde{g}(X^*_{md}, W_{md}^{DVD}, \varepsilon_{md}) \\
\bar{X}_{md} = M_{yr(m),d}^{Theater} X^*_{md}
$$

where the outcome variable $\bar{Y}_{md}$ denotes the sales volume of DVDs/Blu-ray discs sold through three major big-box retailers in DMA $d$ for movie $m$; $M_{yr(m),d}^{DVD}$ is the potential market size for DVD consumers in DMA $d$ in the year of release for movie $m$; latent variable $Y^*_{md}$ can be interpreted as the average number of DVD purchased of movie $m$ for a consumer in market $d$. This average number of DVDs purchased, $Y^*_{md}$, is an unknown smooth function $\tilde{g}$ of latent variable $X^*_{md}$ the average theatrical attendance of movie $m$ per consumer in DMA $d$, $W_{md}^{DVD}$ is a set of explanatory variables for the determinant of DVD purchases, and $\varepsilon_{md}$ is the error term. Although $X^*_{md}$ is not observed, this per-capita theatrical attendance is related to the
observed total theatrical attendance of movie in that market and $M_{yr(m),d}^{Theater}$ the market size of movie-goers in DMA $d$ at the year of release of the movie $m$. Explanatory variables $W_{md}^{DVD}$ comprise the price of the DVD at release, and commonly controlled-for movie characteristics in the literature including but not limited to movie genre, production budget, advertising spending, number of opening screens.

We are interested in estimating the following model for the relationship between theatrical attendance and subsequent DVD/Blu-ray sales, derived by taking logarithm of the system described above. The log-log specification allows us to interpret the estimated causal effect as elasticity, that is, the percentage change in subsequent DVD/Blu-ray sales volume in big-box retailers as a result of the percentage change in theatrical admission.

$$Y_{md} = \log M_{yr(m),d}^{DVD} + g(X_{md}, W_{md}^{DVD}, \epsilon_{md})$$

where

$$Y_{md} = \log \bar{Y}_{md}$$

$$X_{md} = \log \bar{X}_{md}$$

and function $g$ is a transformed $\bar{g}$ and is assumed to be strictly monotonic in the error term $\epsilon_{md}$. In other words, $Y_{md}$ denotes the log of sales volume of DVDs/Blu-ray discs sold through three major big-box retailers in DMA $d$ for movie $m$; and $X_{md}$ denotes the log of movie attendance in DMA $d$ for movie $m$.

The unknown potential market sizes for DVD and theatrical consumption, $\log M_{yr(m),d}^{DVD}$ and $\log M_{yr(m),d}^{Theater}$, are treated as fixed effects in our model. These DMA-year fixed effects capture between-DMA differences so that our analysis can focus on the within-DMA causal effect that has consumer behavioral interpretation. We choose to use DMA-year fixed effects, rather than DMA fixed effects, because the population of geographical markets fluctuate differently over time. For example, the population in the Houston/Sugar Land/Baytown metropolitan area grew 14% from 2005 to 2010, whereas the population in the Detroit/Warren/Livonia metropolitan area shrank by 3% from 2005 to 2010 (U.S. Census Bureau, 2005, 2010).
**Empirical Strategy**

The identification challenge arises from omitted variables, in spite of the inclusion of DMA-year fixed effects and explanatory variables. Omitted-variable bias could arise from unobserved movie popularity factors that our other explanatory variables do not fully capture. More popular movies are likely to have both higher theatrical viewership and higher DVD/Blu-ray sales, thus confounding the causal effect of theatrical viewership on DVD sales. Compounding the omitted variable issue, the unobserved popularity factor can differ across DMAs even for the same movie. For example, films with Christian themes can have broad appeals in cities with larger proportion of Christians, but may not be so popular in cities with smaller proportion of Christians. And the religious composition of cities varies widely: 48% of adults in the San Francisco metropolitan area and 78% of adults in the Dallas metropolitan area identify as Christians (Pew Research Center, 2014 U.S. Religious Landscape Study). Because unobserved popularity factors can vary by movies and by cities, including movie fixed effects in the model would not resolve the omitted variable bias satisfactorily. Econometrically, the existence of unobserved confounding factors that influence both DVD purchase and theatrical attendance decisions means that $X_{md}$, the log-transformed theatrical attendance, is not conditionally independent of $\varepsilon_{md}$, the error term in the determinant of DVD/Blu-ray sales. Not addressing this endogeneity issue would lead to biased estimate of the causal effect of interest.

These unobserved movie popularity factors may enter into DVD sales equation inside the function $g$. Consider the following example of a highly stylized model: suppose the underlying structure is that each DMA has homogenous agents with a DVD purchase decision rule of $g = 1[U(X_{md}, W_{md}^{DV}, \psi_{md}^{*} ; \Theta_{d}) > 0]$ where $1$ is the indicator function, $U$ is the utility of buying DVD for these homogenous agents of type $\Theta_{d}$, and $\psi_{md}^{*}$ is an unobserved movie factor that affects the utility of DVD purchases. The effect of unobserved confounder $\psi_{md}^{*}$ is unexplained by the model using available observables, and thus is part of the econometric error term. Therefore, the error term $\varepsilon_{md}$ would enter the function $g$ in this example.

To overcome these identification challenges, we need a source of plausibly exogenous variation in theatrical attendance that is independent and correlated with these unobserved confounders, conditional on
the explanatory variables. The occurrence of a snowstorm during the theatrical opening weekend is an ideal instrument for theatrical attendance. Occurrence of snowstorm during the theatrical opening weekend that affect theatrical attendance without directly affecting the DVD/Blu-ray sales volume. In other words, snowstorms “move” the theatrical attendance in a way that is conditionally independent from the unobserved confounders. We can then disentangle the true effect of higher theatrical viewership on subsequent DVD/Blu-ray demand from the effects of confounders, by analyzing the change in subsequent DVD/Blu-ray demand as a result of these exogenous changes in theatrical attendance. We explain in the following paragraphs that snowstorms around theatrical openings are suitable instruments for identifying the effect of theatrical attendance on DVD/Blu-ray demand, because (1) snowstorms during theatrical release affect theatrical attendance, (2) occurrences of snowstorms is random conditional on cities and time of year, (3) major snowstorms can be predicted at most seven days in advance (The New York Times, 2016), and thus studios cannot reschedule a theatrical opening date to avoid an upcoming snowstorm in a particular city, and (4) these snowstorms do not have any lingering direct effect on the demand for DVDs/Blu-ray discs released four to five months after the theatrical release.

When snowstorms happen during the opening weekend, movie attendance decreases. This is because snowstorms impede consumers’ travel to theaters and in turn cause some moviegoers to stay home. Moreover, not all of these affected moviegoers would see the missed movie in theaters in later weeks. Our data suggest only about a third of the lost theater attendance is recouped in the weeks subsequent to the opening weekend, and therefore a snowstorm during the opening weekend has a lasting impact on the eight-week aggregate theatrical viewership. In summary, snowstorms significantly influence theatrical attendance in a market.

Snowstorms are random and can only be predicted with short lead-time. Conditional on the calendar month and the city, the occurrence of snowstorms on any given weekend is random. The formation of a snowstorm is forecasted, at most, one to two weeks ahead. Movie studios schedule movie releases several months ahead of the actual opening date, and thus the studios cannot accurately predict whether a snowstorm will occur during a scheduled theatrical opening. Furthermore, studios almost never postpone a
movie’s release after receiving an accurate forecast of a snowstorm, because of the challenge in last-minute schedule negotiations with cinemas and the cost of additional advertising for any new release date in a particular city. The randomness and unpredictability of snowstorms and the high cost of last-minute rescheduling of theatrical release suggest that, the coincidence of a snowstorm on a theatrical opening weekend should be conditionally independent from any unobserved confounders conditional on the explanatory variables.

In addition, a snowstorm’s effect on the impacted cities is transient. Snowstorms in the United States usually last two to five days. Because DVDs are released four to five months after theater releases, the occurrence of a snowstorm at the time of theatrical opening is highly unlikely to directly affect the sales of the DVDs. Any effect of snowstorms on DVD/Blu-ray sales should be attributed to the indirect effect of snowstorms influencing theatrical attendance in the area and in turn the change in theatrical attendance affecting the subsequent DVD/Blu-ray sales. The view that snowstorms have transitory effect on the affected area is supported by Bloesch and Gourio (2015) who analyzed state-level economic time series and concluded that temperature and snowfall shocks have only short-lived economic effects.

Model Specification

Using occurrences of snowstorms as instrument for theatrical attendance, we expand the aforementioned equations system to the following model

\[ Y_{md} = \log M_{d}^{DVD} \cdot m_{d} + g(X_{md}, W_{md}^{DVD}, \varepsilon_{md}) \]

\[ X_{md} = \log M_{d}^{Theater} \cdot m_{d} + h(Z_{md}, W_{md}^{Theater}, \eta_{md}) \]

where \( Y_{md} \) denotes the log of sales volume of DVDs/Blu-ray discs sold through three major big-box retailers in DMA \( d \) for movie \( m \); \( X_{md} \) denotes the log of movie attendance in DMA \( d \) for movie \( m \); \( Z_{md} \) is set of snowstorm instruments; \( M_{d}^{DVD} \cdot m_{d} \) and \( M_{d}^{Theater} \cdot m_{d} \) are the potential market sizes for DVD and theater in DMA \( d \) in the year of release for movie \( m \); \( g \) is an unknown smooth function that is strictly monotonic in the error term \( \varepsilon_{md} \); \( W_{md}^{DVD} \) is a set of explanatory variables for the determinants of DVD purchases; \( h \) is
an unknown smooth functions that is strictly monotonic in $\eta_{md}$, and $\eta_{md}$ is an scalar error term in the equation of the determinants of theatrical attendance; $W_{md}^{Theater}$ is the set of explanatory variables for the determinant of theatrical attendance.

**Object of Interest**

We are interested in the “average partial effect” (APE) (cf. Blundell & Powell (2003); Wooldridge (2005)) of log theatrical attendance $X_{md}$ on log DVD sales volume $Y_{md}$.

$$APE = E \left[ \frac{\partial}{\partial x_{md}} Y_{md}(x_{md}, W_{md}^{DVP}) \right]$$

The average partial derivative of $Y_{md}$ with respect to the endogenous variable $X_{md}$ can be an important measure of the marginal effect of an exogenous shift in the endogenous variable (Blundell & Powell, 2003). The average partial effect is commonly used for the inference of average effect (e.g. Bester & Hansen (2009); Blundell & Powell (2003); Florens, Heckman, Meghir, & Vytlacil, (2008); Imbens & Newey (2009)).

**Identification Assumption**

We use the control function approach to handle the endogeneity issue and estimate the average partial effect. The key assumption in the control function approach for identification is that there exists an estimable control variate $V$ such that $X$ and $\varepsilon$ are independent conditional on $V$. In other words, a control function $C$ of $X$ and $Z$ is a function such that the control variate $V = C(X, Z)$ leads to conditional independence $X \perp \varepsilon \mid V$. Kasy (2010) showed that the control function approach is valid for a triangular system such as ours, when the function $h$ is strictly monotonic in the error term $\eta$ and error terms $\eta$ and $\varepsilon$ are both unidimensional. The control function approach has been used in different settings in the marketing literature (Luan & Sudhir, 2010; Petrin & Train, 2010). Imbens & Newey (2009) showed that $F_{X_{md}|Z_{md}, W_{md}^{Theater}}(x_{md}, z_{md}, W_{md}^{Theater})$, the conditional cumulative distribution function of endogenous variable $X_{md}$ given instrument $Z_{md}$ and explanatory variables $W_{md}^{Theater}$, is a valid control variate for a
triangular model such as ours (we drop the indices for notational clarity). Furthermore, they proved that this control variate enables the identification of the average partial effect, as given by

\[ APE = E \left[ \frac{\partial}{\partial x_{md}} m(x_{md}, v_{md}, w_{md}^{DV}) \right] \]

where \( m(x_{md}, v_{md}, w_{md}^{DV}) = E[Y_{md}|X_{md} = x_{md}, V_{md} = v_{md}, W_{md}^{DV} = w_{md}^{DV}] \).

Under the assumptions in our model, the average partial effect can be point-identified using continuous, discrete, or even binary instruments (D’Haultfœuille & Février, 2015; Torgovitsky, 2015)

**Estimation Approach**

Imbens & Newey (2009) suggested a two-stage approach to estimate the average partial effect. The control variate \( \hat{V}_{md} = \hat{F}_{X_{md}|Z_{md}W_{md}^{Theater}}(x_{md}, z_{md}, w_{md}^{Theater}) \) is estimated in the first stage, and then the function \( m(x_{md}, v_{md}, w_{md}^{DV}) \) is estimated by using the fitted control variate \( \hat{V} \) together with the endogenous variable \( X_{md} \) and explanatory variables \( w_{md}^{DV} \) to estimate the conditional mean function \( \hat{m}(x_{md}, v_{md}, w_{md}^{DV}) \) in the second stage.

The first stage estimates the control variate which requires us to estimate the distribution of the endogenous variable \( X \) conditional on instruments \( Z \) and explanatory variables \( W^{Theater} \). We use kernel method to nonparametrically estimate this conditional distribution \( F_{X|Z,W} \). For the simplifying the exposition, our description of the estimating approach will lump the explanatory variable \( W^{Theater} \) and instruments \( Z \) into a single set of variables \( Z^* \), that is, \( Z^* = (Z, W^{Theater}) \). Lumping the two set of variables is without loss of generality, because a control function does not differentiate between conditioning on an instrument or an explanatory variable.

We use the estimator proposed by Li and Racine (2008) to estimate the conditional cumulative distribution function (CDF) of \( X \) given \( Z^* \),

\[ \hat{F}_{X|Z^*=z} = \frac{1}{N} \sum_{j=1}^{N} \Phi \left( \frac{x - x_j}{b_1} \right) K_0(z_j, z; b_2) / \hat{f}(z) \]
where \( b_1 \) is a positive scalar bandwidth; \( \Phi \left( \frac{x}{b_1} \right) \) is a smooth approximation to the empirical CDF for \( X \); \( K_0 (\cdot; b_2) \) is a generalized product kernel with a vector of positive bandwidths \( b_2 \); \( z_j \) is the \( j \)-th observation of data; \( \hat{f}(z) = \frac{1}{N} \sum_{j=1}^{N} K_0 (z_j, z) \) which is the kernel estimator for the density \( f(z) \).

To estimate the conditional CDF at any given point \( z \), this kernel estimator effectively computes an average of the smoothed empirical CDF of \( X \) using all observations in the original data, weighed by the similarity of each observation to the given point \( z \). The generalized product kernel \( K(\cdot; b_2) \) defines the weights. Because we are conditioning on \( p \) variables (\( p = \) number of instruments + number of explanatory variables), the product kernel \( K(\cdot; b_2) \) operates on vectors of \( p \)-length and the bandwidth vectors are also of \( p \)-length. The generalized product kernel is the product of a series of univariate kernels.

\[
K_0 (z_j, z ; b_2) = \prod_{l=1}^{p} K_l(z_{j,l}, z_l ; b_{2,l})
\]

where \( K_l(\cdot; b_{2,l}) \) is an univariate kernel for the \( l \)-th dimension of \( Z^* \), and \( b_{2,l} \) is the bandwidth associated with this univariate kernel; \( z_{j,l} \) and \( z_l \) are the \( l \)-th dimension of \( z_j \) and \( z \), respectively. We use a second-order Gaussian kernel for a continuous variable and a modified Aitchison-Aitken kernel (Li & Racine, 2003) for a categorical variable.

\[
K_l(z_{j,l}, z_l ; b_{2,l}) = \begin{cases} 
K_l^{(c)}(z_{j,l}, z_l ; b_{2,l}) & \text{if } z_{j,l}, z_l \text{ are continuous} \\
K_l^{(d)}(z_{j,l}, z_l ; b_{2,l}) & \text{if } z_{j,l}, z_l \text{ are categorical}
\end{cases}
\]

\[
K_l^{(c)}(z_{j,l}, z_l ; b_{2,l}) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{(z_{j,l} - z_l)^2}{2 b_{2,l}^2} \right)
\]

\[
K_l^{(d)}(z_{j,l}, z_l ; b_{2,l}) = \begin{cases} 
1 & \text{if } z_{j,l} = z_l \\
\frac{1}{b_{2,l}} & \text{otherwise}
\end{cases}
\]

---

\(^2\) Even though one can construct a conditional CDF estimator using the unsmoothed indicator functions of the dependent variable (\( X \) in this case), Yu and Jones (1998) recommends the smoothed approach.
The choice of bandwidth parameters is crucial in nonparametric method, whereas the choice of kernel is relatively unimportant (DiNardo & Tobias, 2001). We choose the bandwidth parameters using the leave-one-out cross-validation approach in Li, Lin, and Racine (2013). Their approach sidesteps the computational-intensive numerical integration in each iteration of the minimization of the cross-validation objective function. The cross-validation procedure still takes considerable time even with this speed-up, as the run time increases at least quadratically as the number of observations grows. To make the procedure computationally tractable for our data, we apply to cross validation procedure to a random third of the data to find the bandwidths, and then re-scale these bandwidths for the entire sample. Re-scaling the bandwidth is justified theoretically because the asymptotically optimal bandwidth for second-order kernels is $N^{-1/5}$ times an unknown sample-size-invariant constant that depends on the roughness of the true conditional mean function (Wasserman, 2006). This type of approximation, that performs quadratic time-complexity operations on a random subset of data and then applies the results back to the original data, has been used in the kernel methods literature, e.g. Williams and Seeger (2001). In summary, we find $\tilde{b}_1$ and $\tilde{b}_2$ from a random third of the data, and then rescale $\tilde{b}_1$ and $\tilde{b}_2$ up to obtain $b_1$ and $b_2$, and finally apply the rescaled $b_1$ and $b_2$ to the entire dataset to construct the estimator of the conditional CDF. The fitted control variate $\hat{V}_i$ for each observation can then be calculated by applying the constructed conditional CDF estimator to that observation.

The second stage is the estimation of the conditional mean function of the DVD sales given the fitted control variate and explanatory variables. We use a Nadaraya-Watson regression to nonparametrically estimate $m(D^*)$ the conditional mean function (CDF) of $Y$ given $D^*$, where $D^* = (X, \hat{V}, W^{DVD})$

$$m(D^* = d) \equiv E[Y | D^* = d] = \frac{1}{N} \sum_{j=1}^{N} Y_j K_1(d_j, d; b_3) / \hat{f}(d)$$

where $Y_j$ is the $j$-th data point of the dependent variable; $K_1(\cdot; b_3)$ is a generalized product kernel with a vector of positive bandwidths $b_3$; $d_j$ is the $j$-th observation of data; $\hat{f}(d) = \frac{1}{N} \sum_{j=1}^{N} K_1(d_j, d; b_3)$ which is
the kernel estimator for the density $f(d)$. The generalized product kernel $K_1$ is the product of a series of univariate kernels, similar to the definition in the described of the first stage estimation.

The bandwidth parameters for the second stage is chosen by the leave-one-out cross-validation procedure discussed in Li and Racine (2003). Similar to the workaround described in the first stage estimation, we use the bandwidths chosen from a random third of the data to inform the choices of bandwidths for the estimation of the conditional mean function on the whole dataset.

The estimator of the average partial effect is constructed from the estimated second-stage conditional mean function, and averaged over the data sample. The estimator of average partial effect is given by

$$\hat{APE} = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial x} \hat{m}(X_i, \hat{V}_i, W_i^{D|D})$$

and we use numerical differentiation to calculate the partial derivative of the fitted conditional mean function $\hat{m}$ with respect to $x$. More specifically, we use the symmetric difference quotient by approximating $\frac{\partial}{\partial x} \hat{m}(X_i, \hat{V}_i, W_i^{D|D})$ the partial derivative with respect to $x$ evaluated at the $i$-th observation by $\frac{\hat{m}(X_i + \delta, V_i, W_i^{D|D}) - \hat{m}(X_i - \delta, V_i, W_i^{D|D})}{2\delta}$ (Serafin & Wnuk, 1987) and choosing $\delta = 0.001$. We use block bootstrap to derive an estimate of the standard error and a confidence interval for the estimator of the average partial effect. Block bootstrap can be used in data exhibiting dependence (Lahiri, 2003).

6. Results

A. The Effect of Snowstorms on Theatrical Attendance

Snowstorms significantly affect theatrical attendance. We run a generalized additive model regression (Hastie & Tibshirani, 1990) to analyze the effect of snowstorms on the theatrical attendance$^3$.

---

$^3$ Our first-stage estimator models the conditional CDF. Using the estimated conditional CDF to draw inference on the conditional mean is inappropriate, because asymptotically optimal bandwidths for the estimation of conditional CDF are not equal to those for the estimation of conditional mean.
We find a point estimate of 0.049 (standard error = 0.020) on the opening-weekend-snowstorm instrument and 0.042 (standard error = 0.013) on the prior-week-snowstorm instrument after controlling for movie characteristics, release month, release year, and after adjusting for market size differences across DMAs. The coefficients on the snowstorm instruments are significant at 0.01 level. The coefficient on opening-weekend snowstorms indicates that a snowstorm hits a city during the theatrical opening weekend for a movie, the eight-week aggregate theatrical attendance of the movie in that city falls by about 5%. And if a city was hit by a snowstorm during the week before theatrical opening date of a movie, the eight-week aggregate theatrical attendance of that movie in this city rises by about 4%. These evidence that snowstorms have significant impact on theatrical attendance, coupled with prior research that found no long-term economic impact from severe winter events, suggest that we can use snowstorms to separate out the causal effect of theatrical attendance on the DVD/Blu-ray sales volume from confounders.

B. Effect of Theatrical Attendance on DVD Sales

Table 4 shows the estimated average partial effect of log theatrical attendance on log DVD/Blu-ray sales. The dependent variable is the log volume of DVDs/Blu-ray discs sold through three big-box retailers in the first eight weeks after the movie’s DVD release in a DMA. We control for the DVD price at release, movie characteristics, release year, release month, and adjust for market size differences across DMAs.

The estimated average partial effect has a point estimate of 0.271 (standard error = 0.026) on the log theatrical attendance, and the effect is significant at 0.05 level. This finding indicates that higher theatrical attendance leads to significantly more DVDs/Blu-ray discs sold for the same movie in the same market. The coefficient estimate implies a 10% increase (drop) in theatrical attendance induced by snowstorms causes about a 2.7% increase (decrease) in the volume of DVDs/Blu-ray discs sold through big-box retailers in the DVD release window.

The estimated causal effect from our model is significantly different if endogeneity is not adjusted for. Table 4 also shows the estimated effect for a nonlinear model that does not control for the control variate, thus not correcting for endogeneity. The endogeneity-unadjusted model yields an estimated effect
of 0.703 with standard error of 0.023. The significant difference of the estimated effects from the two models, one with endogeneity correction and one without, suggests that accounting for endogeneity in our settings is necessary.

In summary, we find empirical evidence that complementary forces, such as multiple-purchase and word-of-mouth effects, outweigh the cannibalizing force for our research question. Said another way, our results show that, on balance, theatrical consumption complements DVD/Blur-ray sales.

C. Falsification Test

The validity of our empirical approach hinges on the identification assumption for the snowstorm instruments. Part of the assumption is that snowstorms during the opening weekend of a movie’s theatrical release do not have any direct effect on the demand for the DVD released four to five months afterward. We conduct a falsification test to gauge whether this identification assumption holds. The intuition behind our falsification test is that the exclusion restriction assumption implies that snowstorm occurrences would have no association with the DVD/Blur-ray sales for movies whose theatrical attendance was unaffected by snowstorms.

Nine movies in our data were released only in New York City and Los Angeles and then expanded to national release three to four weeks later. Because these movies were not shown in DMAs outside of New York City and Los Angeles for the first three to four weeks of the initial limited release, snowstorm instruments constructed using the initial limited-release date should have no effect on theatrical attendance for DMAs other than New York City and Los Angeles.

Table 7 presents the results of the falsification test. The falsification test regresses log DVD/Blur-ray sales on the snowstorm instruments constructed using the initial limited-release date. The point estimate of the coefficient on the opening-weekend-snowstorm instrument is -0.018 (standard error = 0.030) and the point estimate of the coefficient on the prior-week-snowstorm instrument is 0.009 (standard error = 0.031). These estimates show snowstorm occurrences that do not affect theatrical attendance do not affect
DVD/Blu-ray sales. This finding suggests the absence of a direct effect of snowstorms on DVD/Blu-ray sales, and lends credibility to the identification assumption in our empirical approach.

7. Discussion

Although there is a well-known correlation between a movie’s theatrical revenue and its DVD/Blu-ray revenue, there is no rigorous empirical research analyzing whether increased theatrical sales for a movie are causally related to increased demand in the subsequent DVD/Blu-ray release window. On one hand, an increase in theatrical attendance would substitute for DVD/Blu-ray demand if theatrical experience is relatively undifferentiated from the experience of watching a DVD/Blu-ray at home; on the other hand, an increase in theatrical attendance could boost DVD/Blu-ray demand if two channels are differentiated and/or complementary forces from the multiple-purchases effect and the word-of-mouth effect are large. The direction of the net causal effect of the cannibalization versus complementary forces between these two channels has not been answered in the literature. Understanding the causal relationship between these two channels could be particularly important for the motion picture industry given recent reductions in movie release windows, increases in movie ticket prices, and declines in overall theatrical attendance.

Our research addresses this question by using snowstorms as an exogenous shock to the number of people who see a movie in theaters. Our results demonstrate strong empirical evidence that higher theatrical attendance in a market causes higher DVD/Blu-ray sales in the movie’s subsequent home entertainment release in the same market. Specifically, we estimate that a 10% drop in theatrical attendance induced by snowstorms causes about a 2.7% decrease in the volume of DVDs/Blu-ray discs sold through big-box retailers in the DVD release window.

---

4 The National Association of Theater Owners (NATO) reports that the average release window for movies dropped from 5 months and 22 days in 1998 to 3 months 29 days in 2012 (See Ulin 2013).
5 Time Magazine reports that movie ticket prices hit an all time high in 2014, averaging $8.17 per ticket (Linshi 2015).
6 The Hollywood Reporter reported that the number of people who saw a movie in the theaters hit a two decade low in 2014 (McCintock, 2014).
Although our data do not allow us to identify the mechanism behind the complementarity between these two channels, we conducted a simple online survey that found evidence for each of the mechanisms identified by Hennig-Thureau et al. (2007): the multiple-purchase effect, the informed-cascade effect, and the uninformed-cascade effects.

The insight that the complementary force can outweigh the cannibalization effect of the theatrical channel on the DVD channel has managerial implications for channel optimization. Some marketing actions for the theatrical release window can have a spillover effect on the DVD release window indirectly through changes in theatrical attendance. Therefore, the determination of marketing budgets for each channel should consider not only the relative cost effectiveness and the return on investment of a marketing action on the theatrical channel, but also the spillover these marketing actions have on home entertainment channels.

Our research is, of course, not without limitations. One must exercise caution in extrapolating the magnitude of our finding from a natural experiment to the effects of other theatrical attendance levers on DVD sales. As one of the advantages of natural experiments is the emphasis on internal validity and credibility of the estimand (Imbens, 2010), extrapolating the finding from a natural experiment to other settings can be less credible and requires further study. Another caveat is that our natural experiment identifies the effect of theatrical attendance on demand in the DVD channel but not the other way around. Although the question of reducing the release lag for DVDs is of great interest, our study does not provide direct evidence to answer this question, because an assumption of symmetric complementarity is needed to claim that a marketing-mix lever that makes DVDs a more attractive option would increase theatrical viewership. Furthermore, this paper identifies only the net effect of cannibalization versus complementary forces because we have only aggregate data. Individual-level panel data, possibly augmented with social network records, are needed to tease out the two opposite forces and to separate the relative contribution of various complementary forces.

Our surprising finding that the complementary forces can outweigh cannibalization in theatrical and the downstream DVD retail channel suggests potential avenues for the movie industry to conduct field experiments or further studies to investigate various marketing actions in the theatrical window and to better
understand the mechanisms behind the complementary forces between the theatrical and home entertainment windows.
References


Tables and Figures

Figure 1. Timeline of the announcement of the theatrical opening date, the theatrical window, and the DVD window in the United States.

- Announcement of theatrical opening date: 6-18 months
- Theatrical window: 2-4 months
- DVD window: 4-5 months
Table 1. Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Measure</th>
<th>Source</th>
<th>Level of Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVD volume sold</td>
<td>Total number of DVDs sold for first four weeks of DVD release</td>
<td>movie studios</td>
<td>Movie-DMA</td>
</tr>
<tr>
<td>Box office</td>
<td>Total box office of the first eight-week window from all theaters in the market for the movie</td>
<td>movie studios</td>
<td>Movie-DMA</td>
</tr>
<tr>
<td>Theatrical attendance</td>
<td>Total box office of the first eight-week window from all theaters in the market for the movie divided by the national average movie ticket price in the year of release</td>
<td>Box office: movie studios Average movie ticket price: National Theater Owners Association</td>
<td>Movie-DMA</td>
</tr>
<tr>
<td>Price of DVD at release</td>
<td>Historical price of the DVD on Amazon.com during the first week of DVD release</td>
<td>Camelcamelcamel.com</td>
<td>Movie</td>
</tr>
<tr>
<td>Production budget</td>
<td>Production budget</td>
<td>Internet Movie Database</td>
<td>Movie</td>
</tr>
<tr>
<td>Advertising expenditures</td>
<td>Advertising expenditures in US</td>
<td>Kantor</td>
<td>Movie</td>
</tr>
<tr>
<td>Number of opening theaters</td>
<td>Number of theaters for opening week</td>
<td>Internet Movie Database</td>
<td>Movie</td>
</tr>
<tr>
<td>Movie genre</td>
<td>Movie genre (Action, Comedy, Drama, Family/Animation, Horror)</td>
<td>Internet Movie Database</td>
<td>Movie</td>
</tr>
<tr>
<td>Movie studio</td>
<td>Categorical variable denoting the three movie studios</td>
<td></td>
<td>Movie</td>
</tr>
<tr>
<td>Stars cast indicator</td>
<td>Dummy variable indicating whether this movie has any cast in IMDB’s STARmeter Top 10 list</td>
<td>Internet Movie Database</td>
<td>Movie</td>
</tr>
<tr>
<td>MPAA rating</td>
<td>MPAA rating (G, PG, PG-13, R)</td>
<td>Internet Movie Database</td>
<td>Movie</td>
</tr>
<tr>
<td>IMDB user rating</td>
<td>Review rating for the movie based on average votes by IMDB users</td>
<td>Internet Movie Database</td>
<td>Movie</td>
</tr>
<tr>
<td>Total budget of competing movies in the first week of theatrical release</td>
<td>Sum of the production budgets of movies that were released in the same week as the focal movie</td>
<td>Internet Movie Database</td>
<td>Movie</td>
</tr>
<tr>
<td>Year of theatrical release</td>
<td>The calendar year of the movie opening in theaters</td>
<td>BoxofficeMojo.com</td>
<td>Movie</td>
</tr>
<tr>
<td>Month of theatrical release</td>
<td>The calendar month of the movie opening in theaters</td>
<td>BoxofficeMojo.com</td>
<td>Movie</td>
</tr>
<tr>
<td>Occurrence of any snowstorm during the opening weekend of theatrical release</td>
<td>Dummy variable indicating whether a snowstorm occurred in the DMA at any point during the opening weekend of theatrical release</td>
<td>National Climate Data Center – Storm Event Database</td>
<td>Movie-DMA</td>
</tr>
<tr>
<td>Occurrence of any snowstorm during the 7-day window</td>
<td>Dummy variable indicating whether a snowstorm occurred in the DMA</td>
<td>National Climate Data Center – Storm Event Database</td>
<td>Movie-DMA</td>
</tr>
<tr>
<td>prior to the theatrical release date</td>
<td>during the 7-day window prior to the theatrical release date</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Table 2. Summary Statistics and Correlation Matrix

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) DVD_VOLUME</td>
<td>5844</td>
<td>12453</td>
<td>1</td>
<td>468913</td>
</tr>
<tr>
<td>(2) BOX_OFFICE</td>
<td>407136</td>
<td>1017869</td>
<td>457</td>
<td>27900000</td>
</tr>
<tr>
<td>(3) DVD_RELEASE_PRICE</td>
<td>16.74</td>
<td>3.36</td>
<td>6</td>
<td>27</td>
</tr>
<tr>
<td>(4) OPEN_THEATERS</td>
<td>3055</td>
<td>531</td>
<td>1583</td>
<td>4045</td>
</tr>
<tr>
<td>(5) AD_EXPENSE ($'MM)</td>
<td>29.75</td>
<td>12.77</td>
<td>5.32</td>
<td>62.26</td>
</tr>
<tr>
<td>(6) BUDGET ($'MM)</td>
<td>63.7</td>
<td>49.9</td>
<td>9.0</td>
<td>250.0</td>
</tr>
<tr>
<td>(7) COMPETE_BUDGETS ($'MM)</td>
<td>99</td>
<td>71</td>
<td>35</td>
<td>380</td>
</tr>
<tr>
<td>(8) IMDB_RATING</td>
<td>6.2</td>
<td>0.9</td>
<td>3.5</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correlation</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) DVD_VOLUME</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) BOX_OFFICE</td>
<td>0.73</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) DVD_RELEASE_PRICE</td>
<td>0.00</td>
<td>0.05</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) OPEN_THEATERS</td>
<td>0.25</td>
<td>0.20</td>
<td>0.11</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5) AD_EXPENSE ($'000)</td>
<td>-0.01</td>
<td>0.09</td>
<td>0.13</td>
<td>0.53</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6) BUDGET ($'MM)</td>
<td>0.22</td>
<td>0.18</td>
<td>0.12</td>
<td>0.63</td>
<td>0.37</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7) COMPETE_BUDGETS ($'MM)</td>
<td>0.00</td>
<td>0.03</td>
<td>0.04</td>
<td>0.14</td>
<td>0.23</td>
<td>-0.09</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>(8) IMDB_RATING</td>
<td>0.19</td>
<td>0.14</td>
<td>0.19</td>
<td>0.16</td>
<td>0.27</td>
<td>0.25</td>
<td>0.02</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 3. Effect of Snowstorms on Theatrical Attendance

<table>
<thead>
<tr>
<th></th>
<th>Dependent variable: log theatrical attendance, adjusted for DMA market size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snowstorm occurred during theatrical opening weekend</td>
<td>-0.049*** (0.020)</td>
</tr>
<tr>
<td>Snowstorm occurred within 7 days prior to opening date</td>
<td>0.042*** (0.013)</td>
</tr>
<tr>
<td>Movie characteristics controls</td>
<td>Yes</td>
</tr>
<tr>
<td>Opening month control</td>
<td>Yes</td>
</tr>
<tr>
<td>Opening year control</td>
<td>Yes</td>
</tr>
<tr>
<td>N_movie</td>
<td>104</td>
</tr>
<tr>
<td>N_DMA</td>
<td>198</td>
</tr>
<tr>
<td>N</td>
<td>20460</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses. All regressions are run on the subset of wide release released from November through March in all DMAs. Movie characteristics controls includes total budget of competing movies during theatrical opening, advertising spending, production budget, opening theaters, genre, MPAA category, presence of movie star, IMDB rating.

* p < 0.1; ** p < 0.05; *** p < 0.01.
Table 4. Effect of Theatrical Attendance on DVD Sales. With and Without Endogeneity Correction.

<table>
<thead>
<tr>
<th>Dependent variable: log DVD/Blu-ray sales volume, adjusted for DMA market size</th>
<th>Our model (with endogeneity correction)</th>
<th>Without endogeneity correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Partial Effect estimate:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log theatrical attendance, adjusted for DMA market size (s.e.)</td>
<td>0.271*** (0.026)</td>
<td>0.703*** (0.023)</td>
</tr>
<tr>
<td>Controls:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price of DVD at release</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of weeks between theatrical and DVD releases</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Characteristics of competing DVDs at the week of DVD release</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Movie characteristics</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Month of DVD release</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year of DVD release</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N_{movie}</td>
<td>103</td>
<td>103</td>
</tr>
<tr>
<td>N_{DMA}</td>
<td>209</td>
<td>209</td>
</tr>
<tr>
<td>N</td>
<td>21,034</td>
<td>21,034</td>
</tr>
</tbody>
</table>

Note: Standard errors, clustered at movie level, are in parentheses. All regressions are run on the subset of wide release released from November through March in all DMAs. Movie characteristics controls includes total budget of competing movies during theatrical opening, advertising spending, production budget, opening theaters, genre, MPAA category, presence of movie star, IMDB rating.

* p < 0.1; ** p < 0.05; *** p < 0.01.
**Table 5: Falsification Test of IV Strategy. Do the Instruments Affect DVD Sales Directly?**
Reduced-Form Result for Sample of Limited-Release Movies that Expanded to National Release at Least 2 Weeks after Initial Release

<table>
<thead>
<tr>
<th>Dependent variable: Log DVD Volume</th>
<th>(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limited release, exclude-NY, LA sample</td>
<td></td>
</tr>
<tr>
<td>Opening-weekend-snowstorm indicator</td>
<td>-0.018</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
</tr>
<tr>
<td>Prior-week-snowstorm indicator</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>Yes</td>
</tr>
<tr>
<td>DMA fixed effects</td>
<td>Yes</td>
</tr>
<tr>
<td>N_{Movie}</td>
<td>9</td>
</tr>
<tr>
<td>N_{DMA}</td>
<td>120</td>
</tr>
<tr>
<td>N</td>
<td>1080</td>
</tr>
<tr>
<td>R^2</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Note: Robust standard errors, clustered at movie level, are in parentheses. The regression is run on the limited releases that were released from November through March, and then expanded to national wide release at least two weeks after the initial limited release.

* p < 0.1; ** p < 0.05; *** p < 0.01.
Appendix

A. Survey Questions

1. How many (approximately) movies did you see in movie theaters in the last 5 years?

2. How many (approximately) movie DVDs did you purchase in the last 5 years?

3. In the last 5 years, for what percentage of all the movies you have seen in a movie theater did you later also purchase the DVD?

4. What are your main reasons for buying the DVDs after you saw the movies in theater?
   ✓ To re-watch the movie
   ✓ Gifts for friends and family
   ✓ For your collection
   ✓ Other reasons
   ✓ I never bought those DVDs

5. Of the DVDs you have purchased in the last 5 years, what percentage of those did you buy because you did not see the movie in theaters, but heard from friends or acquaintances the movie was good?

6. Of the DVDs you have purchased in the last 5 years, what percentage of those did you buy because you did not see the movie in theaters, but the movie was a huge box office success?