THE PERFECT STORM:

Using Snowstorms to Analyze the Effect of Theatrical Attendance on the Demand for Subsequently Released DVDs

Patrick Choi‡
pakyan@andrew.cmu.edu

Peter Boatwright‡
pbhb@andrew.cmu.edu

Michael D. Smith†
mds@andrew.cmu.edu

This Version: June 2015

Acknowledgements: This research was conducted as part of Carnegie Mellon University's Initiative for Digital Entertainment Analytics (IDEA), which receives unrestricted (gift) funding from the Motion Picture Association of America. This research was conducted independently without any oversight or editorial control. All findings and errors are entirely our own.

‡ Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213.
† Heinz College, School of Information Systems and Management and Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213.

The Perfect Storm:

Using Snowstorms to Analyze the Effect of Theatrical Attendance on the Demand for Subsequently Released DVDs

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 (IV) approach, 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 rise (drop) in theatrical attendance causes an approximate 8 percent increase (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.

Keywords: *Movie sales, DVD sales, snowstorms, empirical analysis.*

1. Introduction

The marketing environment for motion picture content has changed significantly in recent years. Although 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 over time, both in terms of revenue and consumer interest. For example, theatrical attendance hit a two-decade low in 2014 (McClintock, 2014), and an early 2015 CBS News survey found that, given the choice, 57 percent of Americans prefer to watch movies at home rather than in the theater.[1] The preference for in-home viewing is consistent with revenue figures showing $17.8 billion in total home entertainment spending in 2014,[2] versus $10.4 billion in total theatrical revenue.[3] 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).

In the midst of 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 con-

[1] http://www.cbsnews.com/news/cbs-news-poll-americans-and-the-movies/
[2] http://degonline.org/wp-content/uploads/2015/01/2014_-DEG-Home-Entertainment-Spending-Final-External_1-5-2015.pdf
[3] http://www.boxofficemojo.com/yearly/

sumption 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).

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 increase (decline) in theatrical attendance causes an 8 percent increase (decline) in DVD/Blu-ray demand. This result suggests that there is significant differentiation between these two products and informs marketing practice in this rapidly evolving market.

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 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. 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 have affected the box office and DVD/Blu-ray sales.

Finally, we limit our sample of movies in two additional ways to accommodate our identification strategy, which uses snowstorms as an exogenous shock to the number of consumers who see the movie in theaters. First, we include only movies that opened in theaters from November through March, because snowstorms, the instrument crucial to our empirical strategy, generally occur only in the winter and early spring. Second, because of the monotonicity assumption necessary for our instrument to be valid, we exclude movies that had very wide national openings (top 20th percentile of number of opening theaters). We do this because the effects of snowstorms on theatrical viewership are different between movies in the top 20th percentile of the number of opening theaters and those in the bottom 80th percentiles.

Figure 1a shows the effect of snowstorms during opening weekends on theatrical attendance for movies in the bottom 80th percentile (≤ 3,550), in the 80th-90th percentile (3,550 to 3,675), and in the top 10th percentile (>3,675) of national opening theaters. As this figure shows, opening-weekend snowstorms significantly lower theatrical attendance for the movies with 3,550 or fewer opening screens, but increase the theatrical attendance for the movies with between 3,551 and 3,675 opening theaters, and significantly boost the box office for the movies with more than 3,675 opening theaters. A plausible explanation for why the number of opening theaters moderates a snowstorm's effect on theatrical attendance is that a lower number of national opening theaters implies a generally longer travel distance for consumers to the closest showing theater, and longer travel distance induces heavier-than-normal utility penalties during a snowstorm. Therefore, snowstorms have a strong negative impact on theater attendance for the sparsely screened movies. On the other hand, consumers may not have to travel far to see a movie with a large number of showing theaters, and thus snowstorms can have less of an impact on consumers' decisions to see these widely screened movies.

Furthermore, snowstorms can cause a substitution of demand from sparsely screened movies to widely screened movies. Some consumers who would have gone to a sparsely screened movie in the absence of a snowstorm may switch to see another movie shown in a nearby theater in the presence of a snowstorm. This substitution effect explains the surprising phenomenon of an opening-weekend snowstorm increasing the theatrical attendance of movies with very large numbers of opening theaters (see Figure 1b). The identification of the local average treatment effect in the presence of heterogeneous treatment effects using the IV approach requires each instrument to affect all observations in the same direction. As a result, we exclude from the model the movies with the number of opening theaters above the 80th percentile, which leaves us with 84 movies in our data. The three studios provide data for 204 US designated market areas (DMAs). Our analyses focus on 143 DMAs after we exclude the DMAs that have no snowstorms between 2003 and 2014 and the DMAs that have unreliable or questionable severe-weather data.

The unit of analysis is the outcome of a movie in a DMA. We have a total of 12,012 observations from the combinations between 84 movies and 143 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 et al., 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 pair-wise 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 divide the total box office 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 to the standard of that 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). We did not use this index in our paper because the most recent Hollywood Reporter's star-power ranking was published in 2006, well before our study period. 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 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

5. is the price of the DVD at DVD release in the local market, we do not have data at such granularity. Therefore, as a control, we use the median value of the daily prices over the first two weeks on Amazon.com. We believe the price of the DVD on Amazon.com has a meaningful correlation with the average DVD price at the big-box store where our DVD sales data are collected.

Explanatory variables at geographic-market level:

6. ***Designated market-area fixed effects:*** We include market fixed effects at the DMA-level to absorb movie-invariant market-specific unobserved heterogeneities.

4. Model specification and Empirical Strategy

We are interested in estimating the following reduced-form model for the relationship between theatrical attendance and subsequent DVD/Blu-ray volume sold for each movie i in each DMA j:

$$log\ DVD\ Volume_{ij} = \beta_{ij}\ log\ Theatrical\ Attendance_{ij} + \eta\ log\ DVD\ Price_i + Movie\ Characteristics_i'\Gamma + \alpha_j + \omega_i + \varepsilon_{ij} \quad (1)$$

where the variable $DVD\ Volume_{ij}$ represents the total number of DVDs/Blu-ray discs sold through three major big-box retailers in market j for movie i, and the variable $Theatrical\ Attedance_{ij}$ represents the total theatrical attendance in all theaters in market j for movie i. The coefficient β_{ij} corresponds to the effect of a change in log theatrical attendance on log DVD/Blu-ray discs sold. We allow this effect to be heterogeneous across markets and movies. We control for the price of the DVD and a set of movie characteristics commonly used in the literature. The model includes fixed effects for each market α_j. Furthermore, the model includes ω_i unobserved shocks that are common across markets, and ε_{ij} unobserved idiosyncratic shocks.

We choose a log-log specification because the market size for DVD/Blu-ray sales measured in our data differs from that for theatrical admission. Our measurement of the DVD market covers only three major big-box retailers and is considerably smaller than the measurement of the theater channel, which covers all theaters. The log-log specification allows us to interpret the β_{ij} coefficient as the percentage change in subsequent DVD/Blu-ray sales in big-box retailers as a result of the percentage change in theatrical admission. Thus, a log-log specification requires weaker assumptions than a level specification to extrapolate the measured percentage-change relationship in the DVD/Blu-ray sales through the big-box-retailer channel to the entire DVD channel.

Our model uses DMA fixed effects to control for unobserved heterogeneities that vary across geographic markets but are constant across movies. Examples of market-specific but movie-invariant unobserved heterogeneities are city demographics that were stable in the past decade, such as gender ratio and political affiliation. These market-specific unobserved heterogeneities are likely to be correlated with theatrical attendance, and not controlling for them would cause an endogeneity issue.

The most important identification problem arises from the omitted-variable bias, in spite of the DMA fixed effects and controls. Ommitted-variable bias could arise from unobserved movie popularity factors that control variables do not fully explain. More popular movies are likely to have both higher theatrical viewership and higher DVD/Blu-ray sales. Another possible source of confounders is unobserved time-varying shocks in a market, for example, an ethnicity composition that changes differently over time across cities. In this case, diverging movie preferences across DMAs from changing ethnicity distribution will cause positive correlations in theatrical and DVD demands in each market. Varying economic conditions across DMAs is another example of cofounders. All these scenarios correspond to correlations between the log-transformed theatrical attendance and the error term in the determinant of DVD/Blu-ray sales (equation 1).

To overcome these identification challenges, we need a source of plausibly exogenous variation in theatrical attendance that is not correlated with these unobserved confounders. The occurrence of a snowstorm during the theatrical opening weekend and the occurrence of a snowstorm during the week prior to theatrical release are ideal instruments for theatrical attendance. These two instruments enable a two-stage least squares (2SLS) approach to estimate the following system of equations:

$$log\ DVD\ Volume_{ij} = \beta\ log\ Theatrical\ Attendance_{ij} + \eta\ log\ DVD\ Price_i + Movie\ Characteristics_i{}'\ \Gamma + \alpha_j + \omega_i + \varepsilon_{ij} \quad (2)$$

$$log\ Theatrical\ Attendance_{ij} =$$
$$\theta_1\ Snowstorm_{ij} + \theta_2\ Prior\ Snowstorm_{ij} + \tau\ log\ DVD\ Price_i + Movie\ Characteristics_i'\Psi + \phi_j + w_i + e_{ij} \quad (3)$$

where ϕ_j is the market fixed effects, and w_i and e_{ij} are unobserved demand shocks to box office receipts.

The key to identification using the 2SLS approach is finding factors (i.e., instruments) that affect theatrical attendance without affecting the DVD/Blu-ray sales volume. Such factors "move" the theatrical attendance in a way that is uncorrelated with unobserved cofounders. 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, and (2) 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. The occurrence of major snowstorms during US movie-opening weekends would impede consumers' travel to theaters and in turn cause some moviegoers to stay home. 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 subsequent weeks of movie showings, and therefore a snowstorm during the opening weekend has a lasting impact on the eight-week aggregate theatrical viewership. On the other hand, as noted above, snowstorms during the week before a movie's opening boost the movie's theatrical attendance. These snowstorms prior to theatrical release cause some consumers to postpone their movie consumption, and some of these consumers may instead watch the newly released movie after the storm subsides. In summary, snowstorms significantly influence theatrical attendance in a market.

A key identification assumption is that, conditional on market fixed effects and movie characteristics, the occurrence of snowstorms prior to and during the theatrical opening weekend in a market does not affect the subsequent DVD/Blu-ray sales except through the indirect effect of snowstorms affecting theatrical viewership. In other words, the occurrence of snowstorms must be uncorrelated with the unobserved cofounders, after controlling for other covariates. This assumption is also known as "exclusion restriction."

Snowstorm instruments should satisfy this key assumption because of the randomness and unpredictability of snowstorms. Conditional on calendar month and market, the occurrence of snowstorms on any given weekend is random. The formation of a snowstorm can only be forecasted 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 the new release date. The randomness of snowstorm occurrence and the high cost of rescheduling theatrical release suggest that, conditional on control variables, the coincidence of snowstorm and theatrical release should be uncorrelated with unobserved cofounders.

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.

Beyond the aforementioned identification assumptions, we posit that all movie-market units respond to an instrument in the same direction. This assumption is commonly known as a monotonicity assumption. In the context of this paper, if snowstorms during opening weekends depress theatrical attendance in some movie markets, then no-opening-weekend snowstorm would boost theatrical attendance in any movie market. However, the direction of the effect of opening-weekend snowstorms is not consistent in the full sample of wide-release movies. As discussed in the data section, for movies in top 10th percentile of the number of national opening theaters, snowstorms during the opening weekend significantly increased theatrical attendance. The direction of the effect of snowstorms is opposite to the effect in the subsample of movies in the bottom 80th percentile of the number of national opening theaters. Similarly, for movies in top 10th to 20th percentile of the number of national opening theaters, the point estimate of the effect of an opening-weekend snowstorm is also positive. Estimating the causal effect of theatrical attendance on DVD/Blu-ray sales using the full sample would violate the monotonicity assumption. To deal with this violation of the monotonicity assumption, we analyze only the movies in the bottom 80th percentile of the number of national opening theaters. In the subsample that excludes movies in the top 20th percentile of the number of opening theaters, opening-weekend snowstorms consistently depress box office receipts in all movie markets and thus satisfy the monotonicity assumption. The monotonicity assumption is crucial for interpreting the IV estimate as the local average treatment effect when we assume the effect of theatrical attendance on DVD/Blu-ray sales varies across movies and markets.

With the identification assumptions and monotonicity assumption, the 2SLS estimate of β in equation (2) corresponds to the weighted average of β_{ij} in equation (1), which is the heterogeneous effect of log theatrical attendance on log DVD/Blu-ray sales (Angrist, Graddy, & Imbens, 2000):

$$\beta = \lambda \frac{E\left[\beta_{ij} * \Delta TA_{ij}^{Snowstorm}\right]}{E\left[\Delta TA_{ij}^{Snowstorm}\right]} + (1-\lambda) \frac{E\left[\beta_{ij} * \Delta TA_{ij}^{Prior\ Snowstorm}\right]}{E\left[\Delta TA_{ij}^{Prior\ Snowstorm}\right]} \tag{4}$$

where

$\Delta TA_{ij}^{Snowstorm} \equiv (\log Theatrical\ Attendance_{ij}\ | Snowstorm_{ij} = 1) - (\log Theatrical\ Attendance_{ij}\ | Snowstorm_{ij} = 0)$ is the counterfactual change in log theatrical attendance for movie *i* in DMA *j* when the opening-weekend-snowstorms instrument $Snowstorm$ switches from zero to one; $\Delta TA_{ij}^{Prior\ Snowstorm} \equiv (\log Theatrical\ Attedance_{ij}\ | Prior\ Snowstorm_{ij} =$

$1) - (\log Theatrical\ Attedance_{ij}\ |Prior\ Snowstorm_{ij} = 0)$ is the counterfactual change in log theatrical attendance for movie i in DMA j when the prior-week-snowstorms instrument $Prior\ Snowstorm$ switches from zero to one. $\lambda \in [0,1]$ is the relative strength of the two instruments in affecting theatrical attendance.

The interpretation of the 2SLS estimate of β in terms of the heterogeneous effect of log theatrical attendance on log DVD/Blu-ray sales involves averaging as follows: (1) averaging over the two instruments using the relative instrument strength λ as weights, and (2) for each instrument z, averaging over all movies and markets and weighed by ΔTA_{ij}^{z}, the counterfactual change in log theatrical attendance in that movie-market observation induced by the instrument. The weighted average using relative instrument strength essentially gives more weight to the instrument that is more influential in theatrical attendance. The weighted averaging over the counterfactual change in log theatrical attendance in effect gives more weight to those movie-market observations for which theatrical attendance in the movie-market is more affected by occurrences of snowstorms.

We use the 2SLS estimator for the IV approach because of the theoretical connection between the 2SLS estimator and the weighted average of heterogeneous local effects. A potential concern about using the 2SLS estimator is that it is not efficient when errors are not independent. To address this concern, we repeat our analyses using two efficient estimators, namely, the two-step GMM estimator and the continuously updating GMM estimator (Hansen, Heaton, & Yaron, 1996). Prior research suggests the continuously updating GMM estimator has better finite-sample properties, especially in the presence of weak instruments (Hahn, Hausman, & Kuersteiner, 2004). We find these two estimators yield estimates nearly identical to the 2SLS estimator (comparison shown in Appendix A), thus alleviating the concern that the asymptotic inefficiency of the chosen estimator skews our reported results.

To deal with bias that arises from the forward-looking behavior of consumers, we include the price of the DVD in the first-stage equation, which is the equation that specifies the determinants of theatrical attendance. If the expected prices of the DVDs are low, some consumers may choose to not watch the movie in theaters and instead wait to buy the DVD and watch it at home. Because some consumers may consider attributes of the DVD version when deciding whether to watch a movie in theater, not controlling for the price of the DVD at release in the equation of determinants of theatrical attendance would lead to a biased estimate of the effect of theatrical attendance on DVD/Blu-ray demand if consumers are forward looking and have expectations about movie-specific DVD prices.

We use inference techniques that are robust to weak instruments. Econometricians and marketing researchers are increasingly aware of the issue of weak instruments (Murray, 2006; Rossi, 2014; Stock, Wright, & Yogo, 2002). An IV estimator using instruments that are strongly correlated with the endogenous variable (after removing the effect of control variables) is unbiased and thus fixes the endogeneity problem. On the other hand, an IV estimator using weak instruments is biased toward the OLS estimator; and in the worst case, the IV estimator with weak instruments removes none of the endogeneity bias in the OLS estimator (Staiger & Stock, 1997). Furthermore, the sampling distribution of the IV estimator with weak instruments has much higher variance than that of the OLS estimator (Rossi, 2014). These two issues lead to unreliable inference from the standard technique in IV regressions with weak instruments. To address the issue of unreliable inference in the case of weak instruments, we conduct a weak-instrument-robust hypothesis test on the coefficient of interest (i.e., the effect of theatrical attendance on

DVD/Blu-ray sales volume). The Wald test is robust to weak instruments through testing the parameter and over-identification restriction jointly (Anderson and Rubin, 1949). We also conduct the conditional likelihood ratio test on the key coefficient (Moreira, 2003; Andrews, Moreira, and Stock, 2006), which is an alternative approach to inference in the presence of weak instruments. Furthermore, we derive the weak-instrument-robust 95% confidence set for the effect of theatrical attendance on DVD/Blu-ray sales volume by inverting the Wald test (Finlay & Magnusson, 2009), while allowing for non-spherical error structure due to heteroskedasticity and clustering.

5. Results

A. The Effect of Snowstorms on Box Office Revenue

Table 3 reports the first-stage estimates for the effect of snowstorms on theatrical attendance. The dependent variable is the log theatrical attendance in the first eight weeks of theatrical release from all theaters in a DMA for a movie. Standard errors reported in Table 3 are heteroskedasticity consistent and are clustered at the movie level.

Snowstorms significantly affect theatrical attendance. The first-stage regression of the main specification yields a point estimate of -0.116 (standard error = 0.036) on the opening-weekend-snowstorm instrument, and a point estimate of 0.093 (standard error = 0.031) on the prior-week-snowstorm instrument (Column 2 of Table 3). Both coefficients on the instruments are significant with p-values of less than 0.01. The coefficient on opening-weekend snowstorms indicates that if a city has a snowstorm during the theatrical opening weekend for a movie, the eight-week aggregate theatrical attendance of the movie in that city falls by about 11%. On the other hand, the coefficient on prior-week snowstorms suggests the eight-week theatrical attendance increases by about 9% if the city had a snowstorm during the week prior to the theatrical release.

We find that controlling for movie characteristics improves the precision of the estimate of the effect of snowstorms on theatrical attendance, as evidenced by the comparison across columns 1 and 2 of Table 3. Moreover, the inclusion of market fixed effects is important because differences across geographic markets explain a large proportion of variations in theatrical attendance and the probability of snowstorm occurrences. After controlling for both DMA fixed effects and movie characteristics, the two instruments together have an F-statistics of 8.08 ($p<0.001$). We use the Kleibergen-Paap variant of F-statistics, which is robust to heteroskedasticity and clustered errors[4] (Kleibergen & Paap, 2006). Because the first-stage F-statistic of 8.08 is below the "rule-of-thumb" threshold of 10 for strong instruments (Staiger & Stock, 1997), we apply a weak-instrument-robust inference technique in the second-stage analysis. Despite the F-statistic being slightly below the recommended threshold, we believe the substantial impact of snowstorms on theatrical attendance suggests we can exploit these exogenous impacts on theatrical attendance to identify the effect of theatrical attendance on the DVD/Blu-ray sales volume.

The coefficients on the movie characteristics have the expected signs in the first-stage regression. The point estimate implies a 1% increase in national advertising spending is associated with a 0.36% increase in theatrical attendance; the presence of a star in the cast of the movie is associated with a 17.5% increase in theatrical attendance. However, these associations are not statistically significant. Two significant influencers of theatrical attendance are (1) whether the movie opened during Christmas school holidays and (2) the number of opening theaters. Movies

[4] If instead we assume the errors are i.i.d., the F statistic would be 41.97 and would be higher than the critical value of 19.93 for assessing the strength of instruments to have a 10% maximal IV size distortion with one endogenous variable and two instruments (Stock & Yogo, 2005).

that open around Christmas and New Years have 68% higher theatrical attendance than the winter-release movies that open outside of this two-week period. A 1% increase in the total number of opening theaters in the United States is associated with about a 2% increase in local theatrical attendance. Movies with more opening theaters tend to continue showing in theaters longer, thus translating to a larger-than-proportional increase in the eight-week cumulative theatrical attendance. Several papers suggest the number of opening theaters is endogenous (Elberse & Eliashberg, 2003; Gopinath et al., 2013). To assess the issue of potential endogeneity in the number of opening theaters, we use the number of available theaters in the United States to instrument for the number of opening theaters, an approach that is similar to exploiting variations in theater supply in Gopinath et al. (2013). The key difference is that Gopinath et al. (2013) use theater-supply variations at the market level, whereas we only have theater-supply data at the national level. Using variations in national theater supply as an instrument, we do not find evidence that the number of opening theaters is endogenous. The lack of evidence to support the presence of an endogeneity issue in the number of opening theaters may be attributed to the coarser granularity of our theater-supply data or the fact that this paper excludes movies with very high (> 3550) and very low (< 600) numbers of opening theaters.

B. Effect of Theatrical Attendance on DVD Sales

<u>Main Effect of Theatrical Attendance</u>

Table 4 shows the OLS and IV estimates of the effect of theatrical attendance on 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 of DVD release in a DMA for a movie. Columns 1 and 2 show the IV results, and columns 3 and 4 show the OLS results. Column 2 corresponds to our main specification, in which we use the opening-weekend-snowstorm indicator and the pre-release snowstorm indicator to instrument for the theatrical attendance, and control for movie characteristics and DMA fixed effects. The standard errors reported are heteroskedasticity consistent and are clustered at the movie level.

The IV estimate for our main specification yields a point estimate of 0.836 (standard error = 0.178) on the log theatrical attendance, which has a p-value of less than 0.01. This finding indicates that higher theatrical attendance leads to significantly more DVDs/Blu-ray discs sold for the same movie in the same market. A Hausman test shows the IV estimate is not statistically different from the OLS estimate, and this result suggests the multiple sources of biases likely to be present in the OLS results may be canceling each other out. First, one of the sources of biases in OLS is omitted variables. The estimation of DVD/Blu-ray sales likely contains multiple omitted variables, such as unobserved movie popularity factors, gradual changes in demographics, and economic conditions across DMAs. In a regression with multiple regressors and multiple omitted variables, determining the direction of the bias in the coefficient on the endgeonous variable is challenging in general (see Appendix B for technical details). Second, the measurement error in the theatrical attendance variable causes the OLS estimate to attenuate (bias downward). The measurement error arises because we use box office receipts as a proxy for theatrical attendance, and some theaters charge less for afternoon showings and more for the 3D version of a movie than the standard version.

The coefficient estimate implies a 10% increase (drop) in theatrical attendance induced by snowstorms causes about an 8% increase (decrease) in the volume of DVDs/Blu-ray discs sold through big-box retailers in the DVD release window. The 95% confidence interval for the percentage increase in the subsequent DVD/Blu-ray sales caused by a 10% increase in theatrical attendance is [5%, 11.7%]. As discussed in the previous section, our snowstorm instruments for theatrical attendance may be weak. To confirm that our finding of higher theatrical attendance leading to stronger subsequent DVD/Blu-ray sales is not an artifact of the weak instrument issue, we conduct a weak-instrument-robust variant of a hypothesis test on the coefficient on log theatrical attendance. The Wald test on the coefficient of log theatrical attendance yields a p-value of 0.03, indicating that the effect of theatrical attendance on DVD/Blu-ray sales volume is significantly positive even after accounting for the presence of weak instruments.[5] The weak-instrument-robust 95% confidence interval for the log theatrical attendance is [0.169, 1.370] and implies the weak-instrument-robust 95% confidence interval for the increase in the subsequent DVD/Blu-ray sales volume caused by a 10% increase in box office receipts is [1.8% , 13.4%]. In summary, using the IV method, we discover that higher theatrical attendance leads to more DVDs/Blu-ray discs sold in the same market. In short, the theatrical channel is complementary to the DVD/Blu-ray retail channel.

Not surprisingly, the price of DVDs is an important determinant of DVD/Blu-ray sales, and is significant at the p=0.1 level. A 10% hike in the price of a DVD at release is associated with a 6% drop in the volume of DVDs/Blu-ray discs sold. None of the movie-characteristics control variables are significant in the second stage. In particular, the coefficient of advertising spending prior to theatrical release is not significant, suggesting the direct effect of pre-theatrical-release advertising has dissipated in the four- to five-month gap between the theatrical release and the DVD release, suggesting that advertising in the theatrical window affects the demand in the subsequent DVD/Blu-ray channel only through an indirect effect of advertising on theatrical attendance.

[5] The conditional likelihood ratio test, which is an alternative approach to inference in the presence of weak instruments, produces a conclusion similar to that of the Wald test.

We conduct an overidentification test to assess the coherency of the two snowstorm instruments. Although the overidentification test cannot test the validity of the instruments, it can check whether all instruments identify the same parameter (Parente & Santos Silva, 2012). Running the Hansen test on the main IV specification yields a p-value of 0.79, which is evidence that the two snowstorm instruments produce similar estimates. With this insight in mind, we repeat the IV estimation using only the opening-weekend-snowstorm indicator – the stronger instrument between the two – to mitigate the problem of weak instruments. Indeed, the weak-instruments problem is less severe for the IV regression with only the opening-weekend-snowstorm indicator, as the first-stage F-statistics improved from 8.08 to 9.02. The second stage yields a point estimate of 0.780 (standard error = 0.252) for the coefficient on log theatrical attendance for log DVD/Blu-ray sales. This point estimate from the single-instrument specification does not differ substantially from the point estimate of 0.836 obtained using both instruments. The weak-instrument-robust Wald 95% confidence interval for the coefficient is [0.019, 1.458] in the single-instrument case. This result confirms the finding that higher theatrical attendance leads to more DVDs/Blu-ray discs sold in the same market and that the theatrical channel is complementary to the DVD retail channel. We use the results from two-instrument IV regressions in the rest of the paper for robustness-to-finite-sample issues. The first finite-sample moment of the IV estimator for a single endogenous regressor does not exist if we have fewer than two instruments. In general, the number of existing moments for the IV estimator equals the number of overidentifying restrictions (Kinal, 1980).

Differential Effect of Theatrical Attendance on DVD Sales by Movie Characteristics

In the second main result, we find the effect of theatrical attendance on DVD sales volume varies significantly for movies with different prices of DVDs at release, but is similar for movies with a wide range of other movie characteristics. Table 4 reports the IV estimates for the interaction between theatrical attendance and movie characteristics. The IV approach yields a point estimate of 1.080 (standard error = 0.393) on the interaction term between log theatrical attendance and log price of DVD at release. The coefficient estimate suggests changes in theatrical attendance have a stronger impact on the DVD/Blu-ray sales for movies with more expensive DVDs than movies with cheaper DVDs. A plausible explanation for the differential impact of theatrical attendance on the DVD/Blu-ray sales is that consumers may consider more factors when deciding whether to buy a more expensive DVD. Buying a $25 DVD may require a stronger justification than buying a $10 DVD. Word-of-mouth from friends or a positive experience from watching the movie in the theater can provide the justification a consumer needs to purchase a more expensive DVD. One might postulate an alternative explanation in which a differential impact across certain movie characteristics, such as production budget, presence of stars, or genre, causes the observed differential impact across the price of DVDs, and those movie characteristics correlate with the price of DVDs. However, our finding that theatrical attendance does not have a differential impact on DVD/Blu-ray sales in movies with higher production budgets, more advertising spending, more opening theaters, presence of stars, or higher IMDB movie ratings does not support this alternative explanation (columns 2-6 of Table 5).

Differential Effect of Theatrical Attendance on DVD/Blu-ray Sales by Movie Genre

We do not find statistical evidence that theatrical attendance has a differential effect on DVD/Blu-ray sales across movie genres (Table 6). We find a p-value of 0.47 for the joint hypothesis test that the coefficient estimates on log theatrical attendance are equal for horror movies, family movies/animations, action movies, dramas, and comedies. The point estimates for log theatrical attendance for horror movies, family movies/animations, action movies, dramas, and comedies are 1.426, 0.812, 0.718, 0.388, and 0.152, respectively, although in this case, the coefficient estimates are imprecisely estimated. For all of these coefficients, the weak-instrument-robust 95% confidence interval extends beyond [-10, 10].[6]

C. Falsification Test

The validity of our empirical approach hinges on the identification assumption (also known as the "exclusion restriction") for the snowstorm instruments. The assumption is that snowstorms during the opening weekend of 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 assumption of exclusion restriction implies snowstorm occurrences would have no association with the DVD/Blu-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 all DMAs, excluding New York City and Los Angeles.

Table 7 presents the results of the falsification test. The falsification test regresses log DVD/Blu-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 -

[6] We set a limit of -10 for the minimum and 10 for the maximum in the grid search for inverting the Wald test to derive the confidence interval.

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 validity of the identification assumption in our empirical approach.

6. Robustness Checks

A. Sensitivity to Sample Exclusion by the Number of Opening Theaters

In an earlier section, we discussed why we excluded movies with the number of opening theaters in the top 20th percentile. Our rationale is that our data show that snowstorms boost the theatrical attendance of these movies, and therefore including these movies our in study would violate the monotonicity assumption required to identify the local average treatment effect. In this section we conduct a robustness check on our approach of excluding movies based on the number of opening theaters. In our robustness check, we repeat the IV estimation and vary the number of movies included, whereas the inclusion of movies is based on increasing the number of opening theaters. The chosen cutoff in the earlier section leaves 84 movies in the sample. In this sensitivity analysis, we vary the number of movies from 69 to 104 (all wide releases included) in increments of five movies per test. Figure 2a shows the snowstorms' first-stage F-statistic, which is a measure of the joint significance of the two snowstorm instruments in the regression for explaining variations in log theatrical attendance as a function of expansion of the sample inclusion. The F-statistic on the instruments increases initially as more movies are included, but decreases beyond 89 movies. The reversed effect of snowstorms in movies with a higher number of opening theaters likely causes the decrease in the statistical significance of the instruments. Figure 2b shows the coefficient estimates from the second-stage IV estimation of the effect of log theatrical attendance on log DVD/Blu-ray volume sold. The estimated coefficients are consistent across a wide range of different sample exclusions. The estimated effect drops slightly as the sample inclusion expands beyond 84 movies, and drops further as the sample expands to the entire set of wide-release movies. These results suggest our finding of a positive effect of theatrical attendance on subsequent DVD/Blu-ray sales is robust to a variety of different sample exclusion cutoffs.

B. Sensitivity to Log-Log Specification

This paper chooses a log-log specification to model the relationship between the DVD/Blu-ray volume sold through three major big-box retailers and the theatrical attendance of all theaters in a market. The ratio of the potential market size of DVD retailing through the three big-box retailers to the potential market size of the theatrical channel may differ significantly across geographic markets. As we argued above, the log-log specification is more robust than a level (linear) specification, because it measures the percentage change in DVD/Blu-ray volume sold through the three big-box retailers as a result of a percentage change in theatrical attendance in all theaters.

However, one may wonder if our finding of a positive demand relationship between the theatrical and DVD channels is robust to a level specification. Table 8 shows the IV estimate using the level specification. The significance of the instruments in the first stage of the level specification is weaker than that using the log specification of box office, as evident in panel A. The unsurprising drop in instruments' significance is attributed to a large decrease in precision, because markets have starkly different sizes, and the level specification posits an unrealistic assumption that snowstorms cause the same number of consumers on average to change their movie-going decisions across markets of different sizes. Panel B shows that, despite weaker statistical significance of the instruments in the level specification, the estimated effect of theatrical attendance on DVD/Blu-ray volume sold remains positive. This finding suggests our regression results are robust to alternative functional forms.

C. Sensitivity to Timing Window for Defining Snowstorm Instruments

This paper's key finding that higher theatrical attendance leads to higher DVD/Blu-ray sales is obtained using the occurrence of snowstorms during the opening weekend of theatrical release and the occurrence of snowstorms during the week prior to theatrical release as instruments. Table 9 shows our key finding is robust to alternative definitions of the snowstorm instruments. Column 1 of panels 9A and 9B corresponds to the specification used in the main result; column 2 of panels 9A and 9B corresponds to a three-instrument specification in which an instrument for the occurrence of snowstorms between the 14th and 8th day prior to the theatrical release date is added to the two instruments in the main specification; column 3 of panels 9A and 9B corresponds to another three-instrument specification in which an instrument for the occurrence of snowstorms during the weekdays following the theatrical opening weekend is added; column 4 of panels 9A and 9B corresponds to a four-instrument specification in which an instrument for the occurrence of snowstorms during the second week of theatrical release is included, in addition to the weekday-after-opening-weekend-snowstorm instrument.

The point estimates of the first-stage effect of these additional instruments on theatrical attendance have the expected sign, but these effects are not statistically significant. In the first stage of the IV regressions, the instrument for snowstorms occurring two weeks prior to the theatrical release has a coefficient of 0.026 (standard error = 0.037); the instrument for snowstorms occurring on the weekdays after the opening weekend has a coefficient of -0.027 (standard error = 0.027); the instrument for snowstorms occurring during the second week of theatrical release has a coefficient of -0.051 (standard error = 0.039). These point estimates are reasonable because snowstorms prior to theatrical release can create pent-up demand leading to increased theatrical viewership, and snowstorms after theatrical release impede travels leading to decreased theatrical viewership.

The second-stage results in panel B of Table 9 show that the point estimates for the effect of log theatrical attendance on log DVD/Blu-ray sales volume remain fairly consistent after the addition of these new instruments. The weak-instrument-robust Wald tests on all three new specifications are significant at the 0.1 level, which suggests our key finding that higher theatrical attendance leads to higher DVD/Blu-ray sales is robust to different window lengths for defining the snowstorm instruments.

D. Sensitivity of Estimated Effect of Theatrical Attendance on DVD/Blu-ray Volume Sold to Assumptions of Error Structure and Instrument Strength

Table 10 evaluates how the estimated causal effect varies as we relax the model assumptions of error structure and instrument strength. The first row of Table 7 shows the estimated effect based on the strongest assumption: the errors are independent and identically distributed, and the instruments are strong. Under these strong assumptions, the 95% confidence interval for the increase in DVDs/Blu-ray discs sold caused by a 10% increase in theatrical attendance is [6.4%, 10.2%]. The second row shows the estimated effect when we relax the conditional homoscedastic error assumption to allow heteroskedastic errors. The model with heteroskedastic errors shows a wider confidence interval for the effect of interest. The 95% confidence interval for the drop in DVDs/Blu-ray discs sold caused by a 10% drop in theatrical attendance is [6.5%, 10.1%]. The third row shows the estimated effect when we relax both the conditional homoscedastic error and independent error assumptions. Essentially, we use Huber–White standard errors clustered by movies to account for intra-cluster correlation. Relaxing the independent error assumption widens the 95% confidence interval of the causal effect of interest to [4.9%, 11.8%]. In the three models listed above, we assume that instruments are strong. If we were to relax the strong-instrument assumption in addition to the error-structure assumptions, the 95% confidence interval of the causal effect widens to [1.6%, 13.9%]. We show that the width of the confidence interval depends on the error assumption and instrument-strength assumption. Despite the different widths of the confidence intervals, all four specifications strongly reject a zero effect. In summary, the sensitivity analysis provides reassuring evidence that an increase in theatrical attendance leads to stronger DVD/Blu-ray sales.

7. 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 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 theater and DVD purchase histories for movies (see Appendix C for the list of survey questions). Our survey was conducted through Amazon's Mechanical Turk (n=223). We asked the respondents the number of movies they had seen in theaters and the number of DVDs they had bought 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, in the last five years 70% had purchased DVDs 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 estimate 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 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.

8. 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. To the extent that the theatrical experience is relatively undifferentiated from the experience of watching a DVD/Blu-ray at home, one would expect that increased theatrical sales would substitute for DVD/Blu-ray demand. However, to the extent that these two channels are differentiated, increased theatrical sales could complement DVD/Blu-ray demand. Understanding the causal relationship between these two channels could be particularly important for the motion picture industry given recent reductions in movie release windows,[7] increases in movie ticket prices,[8] and declines in overall theatrical attendance.[9]

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% increase (drop) in theatrical attendance induced by snowstorms causes about an 8% increase (decrease) in the volume of DVDs/Blu-ray discs sold through big-box retailers in the DVD release window.

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. Marketing actions for the theatrical release window 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

[7] 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).
[8] Time Magazine reports that movie ticket prices hit an all time high in 2014, averaging $8.17 per ticket (Linshi 2015)
[9] The Hollywood Reporter reported that the number of people who saw a movie in the theaters hit a two decade low in 2014 (McClintock, 2014).

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. Our estimated local average effect is tied to the snowstorm instruments and the subpopulation whose behaviors are influenced by the instruments (Imbens, 2014). The more similar the characteristics of consumers that change their theatrical viewing decisions by other levers to the characteristics of the consumers influenced by snowstorms, the more valid the extrapolation of the effect of theatrical attendance on the DVD sales estimated using snowstorms to the effect using a different marketing action. 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. Despite the lack of direct evidence to answer the question of how to shorten the waiting period between the theatrical release and the DVD release, our finding that theatrical viewership increases DVD sales is informative to both marketing strategy in the currently evolving movie release windows and to future research analyzing the release lag between the theatrical and home entertainment windows.

References

Ainslie, A., Drèze, X., & Zufryden, F. (2005). Modeling Movie Life Cycles and Market Share. *Marketing Science*. Retrieved from http://pubsonline.informs.org/doi/abs/10.1287/mksc.1040.0106

Anderson, T. W., & Rubin, H. (1949). Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *The Annals of Mathematical Statistics*, *20*(1), 46–63. Retrieved from http://projecteuclid.org/euclid.aoms/1177730090

Andrews, D. W. K., Moreira, M. J., & Stock, J. H. (2006). Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression. *Econometrica*, *74*(3), 715–752. http://doi.org/10.1111/j.1468-0262.2006.00680.x

Angrist, J. D., Graddy, K., & Imbens, G. W. (2000). The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish. *Review of Economic Studies*, *67*(3), 499–527. http://doi.org/10.1111/1467-937X.00141

Basuroy, S., Chatterjee, S., & Ravid, S. A. (2003). How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets. *Journal of Marketing*, *67*(4), 103–117. http://doi.org/10.1509/jmkg.67.4.103.18692

Berry, S., Khwaja, A., Kumar, V., Musalem, A., Wilbur, K. C., Allenby, G., … Mele, A. (2014). Structural models of complementary choices. *Marketing Letters*, *25*(3), 245–256. http://doi.org/10.1007/s11002-014-9309-y

Cohen, J. B. (2009). The Money of Movies. *Back Stage*, 50(11), 7–44.

Digital Entertainment Group. (2014). *DEG Year-End 2013 Home Entertainment Report*. Retrieved from http://degonline.org/news-releases/deg-news/deg-releases-year-end-2013-home-entertainment-report/

Elberse, A., & Eliashberg, J. (2003). Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures. *Marketing Science*, *22*(3), 329–354. http://doi.org/10.1287/mksc.22.3.329.17740

Eliashberg, J., & Shugan, S. M. (1997). Film critics: Influencers or predictors? *Journal of Marketing*, *61*(2), 68–78.

Finlay, K., & Magnusson, L. M. (2009). Implementing weak-instrument robust tests for a general class of instrumental-variables models. *Stata Journal*, *9*(3), 398–421. Retrieved from http://www.stata-journal.com/article.html?article=st0171

Gentzkow, M. (2007). Valuing New Goods in a Model with Complementarity: Online Newspapers. *American Economic Review*, *97*(3), 713–744. http://doi.org/10.1257/aer.97.3.713

Gopinath, S., Chintagunta, P. K., & Venkataraman, S. (2013). Blogs, Advertising, and Local-Market Movie Box Office Performance. *Management Science*, *59*(12), 2635–2654. http://doi.org/10.1287/mnsc.2013.1732

Hahn, J., Hausman, J., & Kuersteiner, G. (2004). Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations. *The Econometrics Journal*, *7*(1), 272–306. http://doi.org/10.1111/j.1368-423X.2004.00131.x

Hansen, L. P., Heaton, J., & Yaron, A. (1996). Finite-Sample Properties of Some Alternative GMM Estimators. *Journal of Business & Economic Statistics*, *14*(3), 262–80. Retrieved from http://econpapers.repec.org/RePEc:bes:jnlbes:v:14:y:1996:i:3:p:262-80

Hennig-Thurau, T., Henning, V., Sattler, H., Eggers, F., & Houston, M. B. (2007). The Last Picture Show? Timing and Order of Movie Distribution Channels. *Journal of Marketing*, *71*(4), 63–83. http://doi.org/10.1509/jmkg.71.4.63

Imbens, G. W. (2014). Instrumental Variables: An Econometrician's Perspective. *Statistical Science*, *29*(3), 323–358. Retrieved from http://projecteuclid.org/euclid.ss/1411437513

Keating, Gina. 2012. Netflixed: The Epic Battle for America's Eyeballs. Penguin Group.

Kinal, T. W. (1980). The Existence of Moments of k-Class Estimators. *Econometrica*, *48*(1), 241–249. http://doi.org/10.2307/1912027

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, *133*(1), 97–126. http://doi.org/10.1016/j.jeconom.2005.02.011

Lehmann, D. R., & Weinberg, C. B. (2000). Sales Through Sequential Distribution Channels: An Application to Movies and Videos. *Journal of Marketing*, *64*(3), 18–33. http://doi.org/10.1509/jmkg.64.3.18.18026

Linshi, Jack. 2015. Average Movie Ticket Price Hit All-Time High in 2014. Time Magazine. January 20. (Available from http://time.com/3675462/movie-ticket-prices/, last accessed May 22, 2015).

Liu, Y. (2006). Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing*, *70*(3), 74–89. http://doi.org/10.1509/jmkg.70.3.74

Luan, Y. J., & Sudhir, K. (2010). Forecasting Marketing-Mix Responsiveness for New Products. *Journal of Marketing Research*, *47*(3), 444–457. http://doi.org/10.1509/jmkr.47.3.444

McClintock, Pamela. 2014. Box Office 2014: Moviegoing Hits Two-Decade Low. The Hollywood Reporter, December 31. (Available from http://www.hollywoodreporter.com/news/box-office-2014-moviegoing-hits-760766, last accessed May 22, 2015).

Moreira, M. J. (2003). A Conditional Likelihood Ratio Test for Structural Models. *Econometrica*, *71*(4), 1027–1048. http://doi.org/10.1111/1468-0262.00438

Mukherjee, A., & Kadiyali, V. (2011). Modeling Multichannel Home Video Demand in the U.S. Motion Picture Industry. *Journal of Marketing Research*, *48*(6), 985–995. http://doi.org/10.1509/jmr.07.0359

Murray, M. P. (2006). Avoiding Invalid Instruments and Coping with Weak Instruments. *Journal of Economic Perspectives*, *20*(4), 111–132. http://doi.org/10.1257/jep.20.4.111

Neelamegham, R., & Chintagunta, P. (1999). A Bayesian Model to Forecast New Product Performance in Domestic and International Markets. *Marketing Science*, *18*(2), 115–136. http://doi.org/10.1287/mksc.18.2.115

Nelson, R. A., & Rutherford, P. (2010). The impact of early DVD release on box office revenues. *Applied Economics Letters*, *17*(7), 689–693. http://doi.org/10.1080/13504850802297913

Parente, P. M. D. C., & Santos Silva, J. M. C. (2012). A cautionary note on tests of overidentifying restrictions. *Economics Letters*, *115*(2), 314–317. http://doi.org/10.1016/j.econlet.2011.12.047

Rossi, P. E. (2014). Invited Paper —Even the Rich Can Make Themselves Poor: A Critical Examination of IV Methods in Marketing Applications. *Marketing Science*, *33*(5), 655–672. http://doi.org/10.1287/mksc.2014.0860

Staiger, D., & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, *65*(3), 557–586. http://doi.org/10.2307/2171753

Stock, J. H., Wright, J. H., & Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics*, *20*(4), 518–529. http://doi.org/10.1198/073500102288618658

Ulin, Jeff. (2013). The Business of Media Distribution: Monetizing Film, TV, and Video Content in an Online World, CRC Press, October 15, 2013, p. 242.

Stock, J. H., & Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (pp. 80–108). Cambridge: Cambridge University Press.

Tables and Figures

Table 1. Data

Variable	Measure	Source	Level of Variation
DVD volume sold	Total number of DVDs sold for first four weeks of DVD release	movie studios	Movie-DMA
Box office	Total box office of the first eight-week window from all theaters in the market for the movie	movie studios	Movie-DMA
Theatrical attendance	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	Box office: movie studios Average movie ticket price: National Theater Owners Association	Movie-DMA
Price of DVD at release	Historical price of the DVD on Amazon.com during the first week of DVD release	Camelcamelcamel.com	Movie
Production budget	Production budget	Internet Movie Database	Movie
Advertising expenditures	Advertising expenditures in US	Kantor	Movie
Number of opening theaters	Number of theaters for opening week	Internet Movie Database	Movie
Movie genre	Movie genre (Action, Comedy, Drama, Family/Animation, Horror)	Internet Movie Database	Movie
Movie studio	Categorical variable denoting the three movie studios		Movie
Stars cast indicator	Dummy variable indicating whether this movie has any cast in IMDB's STARmeter Top 10 list	Internet Movie Database	Movie
MPAA rating	MPAA rating (G, PG, PG-13, R)	Internet Movie Database	Movie
IMDB user rating	Review rating for the movie based on average votes by IMDB users	Internet Movie Database	Movie
Total budget of competing movies in the first week of theatrical release	Sum of the production budgets of movies that were released in the same week as the focal movie	Internet Movie Database	Movie
Year of theatrical release	The calendar year of the movie opening in theaters	BoxofficeMojo.com	Movie
Month of theatrical release	The calendar month of the movie opening in theaters	BoxofficeMojo.com	Movie
Christmas holiday theatrical release indicator	Dummy indicating whether the movie opened between US school Christmas/Holiday holidays (December 23 to January 2)		Movie
Designated market area	Fixed-effects dummies for each DMA		DMA
Occurrence of any snowstorm during the opening weekend of theatrical release	Dummy variable indicating whether a snowstorm occurred in the DMA at any point during the opening weekend of theatrical release	National Climate Data Center – Storm Event Database	Movie-DMA
Occurrence of any snowstorm during the 7-day window prior to the theatrical release date	Dummy variable indicating whether a snowstorm occurred in the DMA during the 7-day window prior to the theatrical release date	National Climate Data Center – Storm Event Database	Movie-DMA

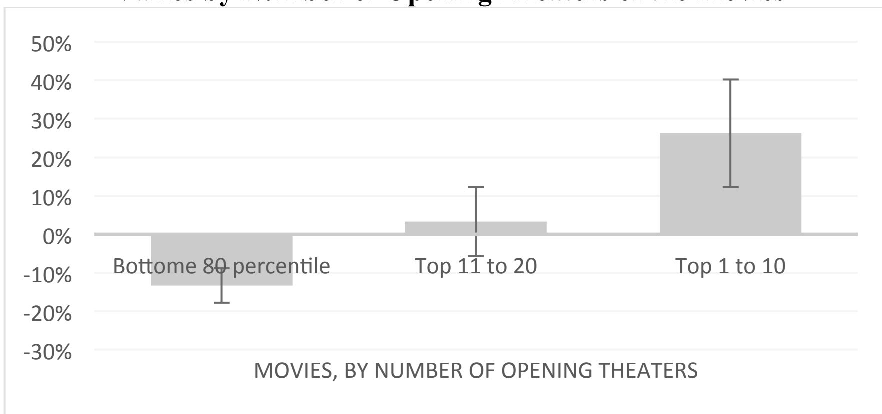
Table 2. Summary Statistics and Correlation Matrix

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

	Correlation							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) DVD_VOLUME	1.00							
(2) BOX_OFFICE	0.73	1.00						
(3) DVD_RELEASE_PRICE	0.00	0.05	1.00					
(4) OPEN_THEATERS	0.25	0.20	0.11	1.00				
(5) AD_EXPENSE ($'000)	-0.01	0.09	0.13	0.53	1.00			
(6) BUDGET ($'MM)	0.22	0.18	0.12	0.63	0.37	1.00		
(7) COMPETE_BUDGETS ($'MM)	0.00	0.03	0.04	0.14	0.23	-0.09	1.00	
(8) IMDB_RATING	0.19	0.14	0.19	0.16	0.27	0.25	0.02	1.00

Figure 1. Effect of Snowstorms on Box Office.

A. Effect of Snowstorm on Opening Weekend on Cumulative Box Office, Varies by Number of Opening Theaters of the Movies



B. Effect of Snowstorm in the Week Prior to Theatrical Release on Cumulative Box Office, Varies by Number of Opening Theaters of the Movies

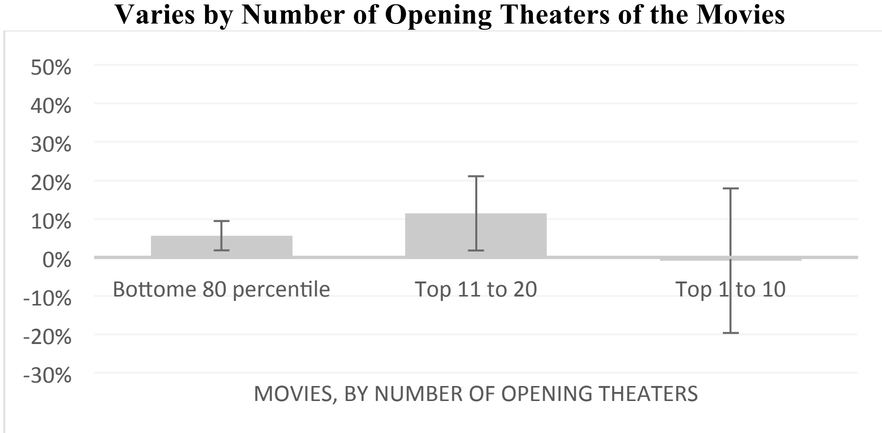


Table 3. Effect of Snowstorms on Box Office. First stage of IV.

	Dependent variable: Log BOX_OFFICE	
	(1)	(2)
	Screens ≤3550 Sample	Screens ≤3550 Sample
Opening Weekend Snowstorm dummy	-0.122* (0.070)	-0.116*** (0.036)
Prior Week Snowstorm dummy	0.072 (0.056)	0.093*** (0.031)
Log OPEN_THEATERS		2.255*** (0.421)
Log AD_EXPENSE		0.365 (0.232)
Log BUDGET		0.039 (0.114)
Log COMPETE_BUDGETS		-0.021 (0.080)
Log DVD_RELEASE_PRICE		-0.151 (0.263)
STARS dummy		0.175 (0.131)
HOLIDAY_OPEN dummy		0.681*** (0.232)
DMA fixed effects	Yes	Yes
Controls	No	Yes
N_{Movie}	84	84
N_{DMA}	143	143
N	12012	12012
R2	0.76	0.89
Kleibergen-Paap F statistic	1.79	8.08
Pr > F	p=0.17	p<0.001

Note: Robust standard errors, clustered at movie level, are in parentheses. All regressions are run on the subset of wide releases released from November through March, where the movies in the top 20th percentile of the number of opening theaters are excluded to maintain monotonicity assumption.
* p < 0.1; ** p < 0.05; *** p < 0.01.

Table 4. Effect of Theatrical Attendance on DVD Sales. IV and OLS estimates.

	DV: Log DVD Volume				
	IV			OLS	
	(1)	(2)		(3)	(4)
	Screens ≤3550 Sample	Screens ≤3550 Sample		Screens ≤3550 Sample	Screens ≤3550 Sample
Log BOX_OFFICE	0.770*** (0.250)	0.836*** (0.178)		0.929*** (0.089)	0.791*** (0.074)
Anderson-Rubin Wald test	*p=0.32*	*p=0.03*			
Anderson-Rubin 95% C.I.		*[0.169 , 1.370]*			
Log_DVD_release_price		-0.600* (0.324)			-0.606* (0.321)
Log OPEN THEATERS		0.635 (0.651)			0.752* (0.445)
Log_ad		0.172 (0.188)			0.188 (0.181)
Log_budget		-0.009 (0.086)			-0.007 (0.086)
Log_compete_budgets		-0.054 (0.085)			-0.055 (0.080)
Stars dummy		-0.193 (0.151)			-0.186 (0.149)
Holiday_open dummy		0.419 (0.271)			0.449* (0.235)
DMA fixed effects	Yes	Yes		Yes	Yes
Controls	No	Yes		No	Yes
N_{Movie}	84	84		84	84
N_{DMA}	143	143		143	143
N	12012	12012		12012	12012
Hansen J test p-value	0.64	0.79			
R2				0.85	0.90
Endogeneity test of Log_box_office p-value	0.64	0.81			

Note: IV regressions of (1) and (2) corresponding to first-stage regressions (1) and (2) in Table 3, respectively. Robust standard errors, clustered at movie level, are in parentheses. All regressions are run on the subset of wide release released from November through March, where the movies in the top 20th percentile of the number of opening theaters are excluded to maintain monotonicity assumption.

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

Table 5. Interaction between Theatrical Attendance and Movie Characteristics; IV Estimates

	DV: Log DVD Volume					
	(1)	(2)	(3)	(4)	(5)	(6)
log theatrical attendance	-1.953* (1.059)	0.755 (3.137)	3.081 (1.736)	4.532 (6.119)	0.822*** (0.170)	2.890 (1.838)
log theatrical attendance x log price of DVD at release	1.080*** (0.393)					
log theatrical attendance x log production budget		0.006 (0.174)				
log theatrical attendance x log advertising expense			-0.223 (0.172)			
log theatrical attendance x log number of opening theaters				-0.454 (0.760)		
log theatrical attendance x stars indicator					-0.301 (0.392)	
log theatrical attendance x IMDB movie rating						-0.350 (0.330)
Movie characteristics	Yes	Yes	Yes	Yes	Yes	Yes
DMA fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N_{Movie}	84	84	84	84	84	84
N_{DMA}	143	143	143	143	143	143

Note: Robust standard errors, clustered at movie level, are in parentheses. All regressions are run on the subset of wide releases released from November through March, where the movies in the top 20th percentile of the number of opening theaters are excluded to maintain monotonicity assumption.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 6. Effect of Theatrical Attendance on DVD Sales by Movie Genre; IV Estimates

	DV: Log DVD Volume
Coefficient for log theatrical attendance	
Horror genre	1.426** (0.677)
Wald 95% confidence set	[<-10 , >10]†
Family/Animation genre	0.812*** (0.272)
Wald 95% confidence set	[<-10 , >10]†
Action genre	0.718** (0.329)
Wald 95% confidence set	[<-10 , >10]†
Drama genre	0.388 (0.585)
Wald 95% confidence set	[<-10 , >10]†
Comedy genre	0.152 (0.468)
Wald 95% confidence set	[<-10 , >10]†
Movie characteristics	Yes
DMA fixed effects	Yes
Controls	Yes
N_{Movie}	84
N_{DMA}	143
N	12012
p-value for joint test of equal coefficients	0.47

Note: Robust standard errors, clustered at movie level, are in parentheses. All regressions are run on the subset of wide releases released from November through March, where the movies in the top 20th percentile of the number of opening theaters are excluded to maintain monotonicity assumption.

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

†: The weak-instrument-robust Wald 95% confidence set extends beyond [-10, 10]. We set a limit of -10 for the minimum and 10 for the maximum in the grid search for inverting the Wald test to derive the confidence interval.

Table 7: Falsification Test of IV Strategy. Do the Instruments Affect DVD Sales Directly?

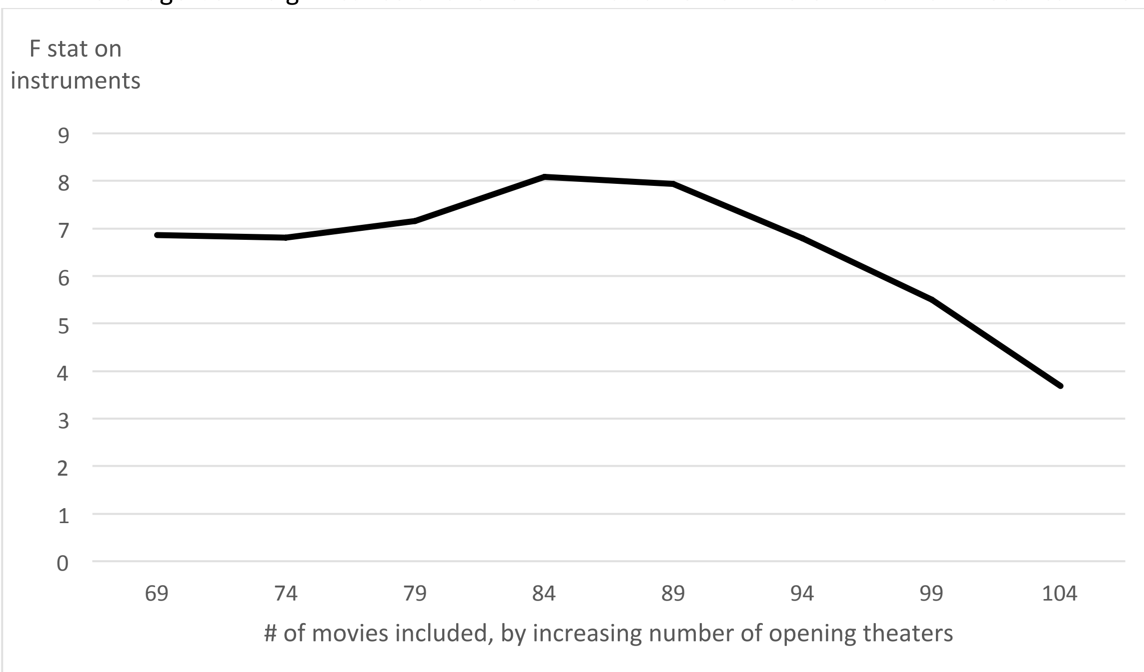
Reduced-Form Result for Sample of Limited-Release Movies that
Expanded to National Release at Least 2 Weeks after Initial Release

Dependent variable: Log DVD Volume	
	(1)
	Limited release, ex-NY, LA sample
Opening-weekend-snowstorm indicator	-0.018 (0.030)
Prior-week-snowstorm indicator	0.009 (0.031)
Year fixed effects	Yes
DMA fixed effects	Yes
N_{Movie}	9
N_{DMA}	120
N	1080
R^2	0.97

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

Figure 2. Robustness Check:
Sensitivity to Sample Exclusion on Number of Opening Theaters

A. First Stage. Joint Significance of Snowstorm Instruments in Determinant of Theatrical Attendance



B. Second Stage of IV. Effect of Theatrical Attendance on DVD Sales

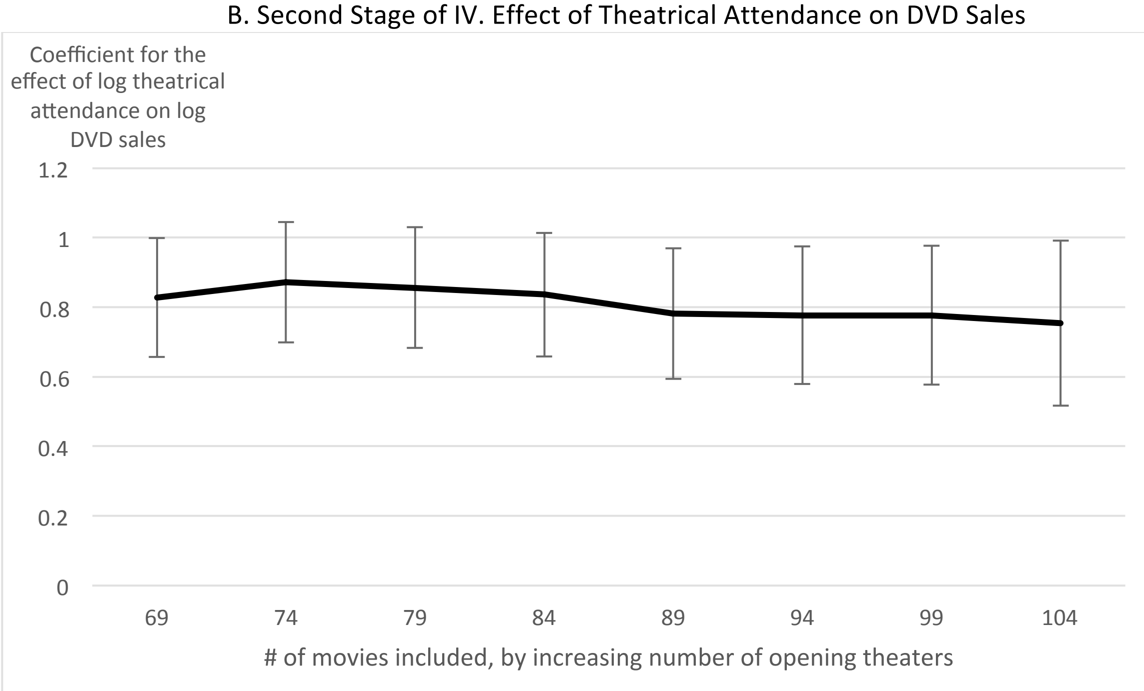


Table 8. Robustness Check: Level Specification

Panel A. Effect of Snowstorms on Theatrical Attendance; First stage of IV

	DV: Theatrical attendance
	(1)
Opening Weekend Snowstorm indicator	-29767.56 (21645.13)
Prior Week Snowstorm indicator	30181.57** (12973.49)
DMA fixed effects	Yes
Controls	Yes
N_{Movie}	84
N_{DMA}	143
N	12012
F statistics	2.96
F test p-value	0.05
R2	0.67

Note: Robust standard errors, clustered at movie level, are in parentheses. All regressions are run on the subset of wide releases released from November through March, where the movies in the top 20th percentile of the number of opening theaters are excluded to maintain monotonicity assumption.

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

Panel B. Effect of Theatrical Attendance on DVD Sales; IV Estimate

	DV: DVD Volume
	(1)
Theatrical attendance	0.019*** (0.004)
DMA fixed effects	Yes
Controls	Yes
N_{Movie}	84
N_{DMA}	143
N	12012
Hansen J test p-value	0.92
Endogeneity test p-value	0.01

Note: Robust standard errors, clustered at movie level, are in parentheses. All regressions are run on the subset of wide releases released from November through March, where the movies in the top 20th percentile of the number of opening theaters are excluded to maintain monotonicity assumption.

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

Table 9. Robustness Check: Sensitivity to Timing Window for Defining Snowstorm Instruments

Panel A. Effect of Snowstorms on Theatrical Attendance; First stage of IV

	Dependent variable: log theatrical attendance			
	(1)	(2)	(3)	(4)
Opening-weekend-snowstorm indicator	-0.116*** (0.036)	-0.117*** (0.037)	-0.114*** (0.036)	-0.111*** (0.036)
Weekday following opening-weekend-snowstorm indicator			-0.027 (0.027)	-0.025 (0.027)
Second-week-snowstorm indicator				-0.051 (0.039)
Prior-week-snowstorm indicator	0.093*** (0.031)	0.091*** (0.031)	0.094*** (0.031)	0.096*** (0.031)
Two-week-prior-snowstorm indicator		0.026 (0.037)		
log number of opening theaters	2.255*** (0.421)	2.624*** (0.421)	2.625*** (0.421)	2.623*** (0.419)
log advertising expense	0.365 (0.232)	0.365 (0.232)	0.365 (0.232)	0.365 (0.231)
log production budget	0.039 (0.114)	0.039 (0.114)	0.038 (0.114)	0.039 (0.114)
log total production budget of competitors in theatrical opening	-0.021 (0.080)	-0.022 (0.080)	-0.022 (0.080)	-0.021 (0.079)
log price of DVD at release	-0.151 (0.263)	-0.153 (0.264)	-0.150 (0.264)	-0.151 (0.262)
Stars indicator	0.175 (0.131)	0.174 (0.131)	0.176 (0.131)	0.172 (0.130)
Holiday opening indicator	0.681*** (0.232)	0.682*** (0.232)	0.679*** (0.233)	0.670*** (0.231)
DMA fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
N_{Movie}	84	84	84	84
N_{DMA}	143	143	143	143
N	12012	12012	12012	12012
R2	0.89	0.89	0.89	0.89
Kleibergen-Paap F statistic	8.08	5.44	5.45	4.68
F test p-value	p<0.001	p=0.002	p=0.002	p=0.002

Note: Robust standard errors, clustered at movie level, are in parentheses. All regressions are run on the subset of wide releases released from November through March, where the movies in the top 20th percentile of the number of opening theaters are excluded to maintain monotonicity assumption.
* p < 0.1; ** p < 0.05; *** p < 0.01.

Panel B. Effect of Theatrical Attendance on DVD Sales; IV Estimates

	DV: Log DVD Volume			
	(1)	(2)	(3)	(4)
log theatrical attendance	0.836*** (0.178)	0.795*** (0.171)	0.789*** (0.183)	1.073*** (0.173)
Wald test	*p=0.03*	*p=0.06*	*p=0.04*	*p=0.02*
Wald 95% C.I.	*[0.169 , 1.370]*	*[-0.083 , 1.150]*	*[0.053, 1.366]*	*[0.435, 1.295]*
log price of DVD at release	-0.600* (0.324)	-0.606* (0.330)	-0.607* (0.334)	-0.560* (0.282)
log number of opening theaters	0.635 (0.651)	0.740 (0.637)	0.757 (0.669)	0.013 (0.614)
log advertising expense	0.172 (0.188)	0.187 (0.186)	0.189 (0.190)	0.085 (0.194)
log production budget	-0.009 (0.086)	-0.007 (0.087)	-0.007 (0.079)	-0.018 (0.081)
log total production budget of competitors in theatrical opening	-0.054 (0.085)	-0.055 (0.081)	-0.055 (0.081)	-0.049 (0.080)
Stars indicator	-0.193 (0.151)	-0.186 (0.152)	-0.185 (0.152)	-0.233 (0.156)
Holiday opening indicator	0.419 (0.271)	0.446* (0.269)	0.450* (0.273)	0.259 (0.253)
DMA fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
N_{Movie}	84	84	84	84
N_{DMA}	143	143	143	143
N	12012	12012	12012	12012
Hansen J test p-value	0.67	0.64	0.40	0.06
Endogeneity test of Log_box_office p-value	0.81	0.99	0.86	0.16

Note: IV regressions of (1) to (4) corresponding to first-stage regressions (1) to (4) in Panel A of Table 9, respectively. Robust standard errors, clustered at movie level, are in parentheses. All regressions are run on the subset of wide releases released from November through March, where the movies in the top 20th percentile of the number of opening theaters are excluded to maintain monotonicity assumption.
* p < 0.1; ** p < 0.05; *** p < 0.01.

Table 10: Sensitivity of Estimated Effect of Theatrical Attendance on DVD Volume Sold to Assumptions of Error Structure and Instrument Strength

Assumptions			Percentage increase in DVDs sold given a 10% increase in theatrical attendance, 95% C.I.
Homoscedastic errors	Independent errors	Strong instruments	
Yes	Yes	Yes	[6.4% , 10.2%]
No	Yes	Yes	[6.5% , 10.1%]
No	No	Yes	[5.0% , 11.7%]
No	No	No	[1.8% , 13.4%]

Appendix

A. Robustness Check: Sensitivity of the Effect of Theatrical Attendance on DVD Sales to the Choice of IV Estimator

	DV: Log DVD Volume		
	2SLS	Two-Step GMM	Continuously Updating GMM
	(1)	(2)	(3)
log theatrical attendance	0.836*** (0.178)	0.832*** (0.178)	0.832*** (0.178)
log price of DVD at release	-0.600* (0.324)	-0.597* (0.324)	-0.599* (0.322)
log number of opening theaters	0.635 (0.651)	0.625 (0.650)	0.626 (0.650)
log advertising expense	0.172 (0.188)	0.173 (0.188)	0.174 (0.189)
log production budget	-0.009 (0.086)	-0.010 (0.085)	-0.011 (0.085)
log total production budget of competitors in theatrical opening	-0.054 (0.085)	-0.058 (0.078)	-0.059 (0.078)
Stars indicator	-0.193 (0.151)	-0.180 (0.142)	-0.180 (0.142)
Holiday opening indicator	0.419 (0.271)	0.405 (0.266)	0.404 (0.267)
DMA fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
N_{Movie}	84	84	84
N_{DMA}	143	143	143
N	12012	12012	12012

B. Bias of OLS Estimate from Multiple Omitted Variables with Multiple Regressors

Suppose we have one endogenous variable X_1, p_2 number of exogenous control variables X_2, and k omitted variables U that have a vector of effects θ on the dependent variable. Without loss of generality, we assume all regressors have zero means. Let the true specification be

$$Y = X_1\beta_1 + X_2\beta_2 + U\theta + \varepsilon.$$

To focus the exposition on the bias of estimator β_1 for the endogenous variable, we partial out the exogenous control variables using the Frisch–Waugh–Lovell theorem:

$$M_{X_2}Y = M_{X_2}X_1\beta_1 + M_{X_2}U\theta + M_{X_2}\varepsilon,$$

where $M_{X_2} = I - X_2(X_2'X_2)^{-1}X_2'$, which is a projection matrix that projects onto the complement of the column space of exogenous control variables X_2.

The OLS estimator of β_1 is $\left(X_1'M_{X_2}X_1\right)^{-1}X_1'M_{X_2}Y$. The bias of the OLS estimator is

$$\begin{aligned} E\left[\hat{\beta}_1^{OLS}\right] - \beta_1 &= plim\left(X_1'M_{X_2}X_1\right)^{-1}X_1'M_{X_2}U\theta \\ &= plim\left(\tilde{X}_1'\tilde{X}_1\right)^{-1}\tilde{X}_1'\tilde{U}\theta \qquad \text{, where } \tilde{X}_1 = M_{X_2}X_1 \text{ and } \tilde{U} = M_{X_2}U. \end{aligned}$$

In essence, the bias of the OLS estimator is a function of $\tilde{X}_1$, the information in the endogenous variable not explained by the exogenous control variables, $\tilde{U}$, the information in omitted variables not explained by the exogenous control variables, and γ, the effects of omitted variables on the dependent variable.

The direction of the bias of the OLS estimator with a single endogenous variable is solely determined by $\tilde{X}_1'\tilde{U}\theta$, because $\tilde{X}_1'\tilde{X}_1$ converges to the variance of a random variable. Decomposing $\tilde{X}_1'\tilde{U}\theta$ as the inner product between $[\tilde{X}_1'\tilde{U}_1 \quad \cdots \quad \tilde{X}_1'\tilde{U}_k]$ and θ shows that the bias of the OLS estimator is a sum of k terms of the form $\tilde{X}_1'\tilde{U}_i\theta_i$. Each of these $\tilde{X}_1'\tilde{U}_i\theta_i$ terms can be thought of as multiplying the effect of one the omitted variables on the dependent variable to the covariance between the portion of the endogenous variable unexplained by the controls and the portion of that omitted variable unexplained by the controls. In other words, the bias can be rewritten as

$$E\left[\hat{\beta}_1^{OLS}\right] - \beta_1 = C_0 \sum_{i=1}^{k} cov\left(\tilde{X}_1', \tilde{U}_i\right) \theta_i \qquad \text{, where } C_o \in [0, \infty).$$

Determining the direction (i.e., the sign) of the bias is hard because (1) we need to determine for each omitted variable the sign of the covariance between the endogenous variable and the omitted variable after we partial out the control variables, and (2) we cannot ascertain the sign of the bias unless the bias contributions from all conceivable omitted variables have the same signs or one of the omitted variables is known to have an overwhelming effect on the dependent variable. Also, the rewritten form of the bias of the OLS estimator clearly shows the bias introduced by each of the omitted variables might cancel each other out.

C. 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?

   ... the movie was a huge box office success?