Product Similarity Network in the Motion Picture Industry

Yanhao Wei*
University of Pennsylvania

This draft: October 14, 2015
For the latest draft, visit:
https://sites.google.com/site/yanhaomaxwei/research

Abstract  This paper studies product entry in the presence of firm learning from the market outcomes of past products. Focusing on the U.S. motion picture industry, we construct a network capturing the similarity amongst the movies released in the last decades. We develop and estimate a model of how the network evolves. Risk-averse firms make go/no go decisions on candidate products that arrive over time and can be either novel or similar to various previous products. We demonstrate that learning is an important factor in entry decisions and provide insights on the innovation vs. imitation tradeoff. In particular, we find that one firm benefits substantially from the learning of the other firms. We find that big-budget movies benefit more from imitation, but small-budget movies favor novelty. This leads to interesting market dynamics that cannot be produced by a model without learning.

*Email: yanhao@sas.upenn.edu. I am indebted to the members of my doctoral committee, Holger Sieg, Eric Bradlow, Joseph Harrington, Katja Seim and Christophe Van den Bulte, for their support on this research. For their helpful suggestions, and also not blaming them for any mistakes in the paper, I want to thank Ron Berman, Bryan Bollinger, Jehoshua Eliashberg, Peter Fader, Hannming Fang, Xiang Fang, Jian Ni, Devin Reilly, Hongxun Ruan, Jagmohan Raju, James Roberts, Andrew Sweeting, Francisco Silva, Qiaowei Shen, Petra Todd, Rakesh Vohra, Pinar Yildirim and Weilong Zhang. I also would like to thank the participants at the 37th Marketing Science Conference and various seminars at the Department of Economics and the Wharton School of the University of Pennsylvania for their comments.
1 Introduction

In many industries, new products roll out at a fast pace, and firms need to constantly anticipate the consumer demand for potential products and make go/no go decisions: Is this prototype going to be make a well-received product? Should I introduce a novel product or imitate some existing products? Examples include motion pictures, book publishing, video games, TV shows, smart-phone apps, cell phone manufacturing, apparel, and even scientific research. Although forecasting the success of a new product bears lots of uncertainty, much can be learned from the market outcomes of past similar products. In this sense, while firms decide what products to introduce, these products in turn affect the product choices of the firms.

This paper focuses on the U.S. motion picture industry to study firm learning from previous products. The industry spends billions of dollars per year, but nevertheless is characterized by a high level of uncertainty on the return of investment (ROI). It is fairly easy to come up with examples like E.T. the Extra-Terrestrial that grossed $360m domestically on an $11m budget, or The Golden Compass that lost $110m from a $180m budget, pushing the studio into bankruptcy. The uncertainty makes imitation a particularly useful strategy. “There continues to be no magic formula for a commercial movie, but patterns emerge, emulating prior successes.”

In fact, movie imitation frequents the media as a subject of discussion as well as debate.

To better understand how firms balance innovation and imitation, we develop a model that focuses on studio’s green-light decisions. A movie’s market outcome is determined by consumer demand over its characteristics such as ideology, storyline, narrative techniques, acting, graphics, music, etc, most of which are unobserved in the sense that they are very difficult to quantify. In empirical entry models, it is often assumed that the unobserved effects are independent across products. In this paper, correlation is explicitly modeled. We capture the correlation structure with a network. In general, the demand for two movies are correlated when they are similar in characteristics, in which case links in the network represent similarity between products.

Given the correlation, a production company is then able to form belief on the demand for a candidate movie by looking at the market outcomes of the released movies. A movie typically takes more than one year to produce. So at any time, each firm holds a portfolio of movies that are in production. We assume that firms can be risk averse and seek to maximize the risk-adjusted profits of the portfolio. To model the supply of candidate products, we apply ideas from the literature on evolving random networks (Newman (2003), Jackson (2010)) and specify a stochastic process where candidates continuously arrive and “attach” to the network at the time. In this sense, each candidate is a creative combination of existing movies. The stochastic nature of the process means that the candidate can be similar to few or many existing movies, offering

---

2 Movie imitation has also been a subject of discussion in media. For example, see “Hollywood Learns Originality Does Not Pay.” May 29, 2015, Financial Times; “Are Blockbusters Destroying the Movies?” New York Times, Jan. 6, 2015.
opportunities of both innovation and imitation.

We bring the model to data. To construct the network, we look at what two movies tend to be liked by the same consumers. In such cases, the knowledge of a high demand for one movie entails a high expectation of the demand for the other. We make use of the item-based collaborative filtering (Desrosiers and Karypis (2011)) that calculates similarity between items based on people’s ratings or purchases. It is known as “People Who Liked This Also Liked” on IMDb.com and “Customers Who Watched This Also Watched” on Amazon.com. We construct a network of nearly 4,500 movies released in the U.S. in the last decades. Through reduced-form analysis, we find that previous similar movies are much more predictive of a movie’s market outcome than the covariates commonly used in movie studies (e.g., budget, genre, star power). We also find evidence that suggests firm learning and risk aversion.

The paper proceeds to estimate the model with the method of simulated moments and conduct counter-factual experiments. Several insights are derived. First, we show that learning is an important factor in the industry. For the movies in the data, learning reduces firm’s uncertainty by over 60 percent, on average. Learning allows a firm to produce big-budget movies, which involve higher risks than their small-budget counterparts. Learning also helps a firm maintain a high level of quality on its movies. We find that a firm has some monopoly power in imitating its own movies, which is a significant barrier against social learning, i.e., learning across firms. Nevertheless, we find that learning has substantial spillover to other firms. For a major studio, the indirect benefits from the learning by the other firms are comparable to the direct benefits from its own learning.

Other insights pertain to the balance of innovation versus imitation. We find that whether to imitate or innovate crucially depends on the investment size. Big-budget movies heavily rely on imitation as a way to reduce risks. However, small-budget movies favor novelty. This is because there is tight lower bound on how much they can lose and a higher level of uncertainty increases the chance for them to make a big hit. In a related counter-factual, we find that a lower level of risk aversion can actually increase the overall level of imitation. One cause is that it allows the production of bigger-budget movies where imitation is still necessary. These results provide some unique insights to the rise of blockbusters and the debate surrounding it.\(^2\)

In terms of general insights, this paper adds to the studies on product networks. Compared with the widespread attention on social networks, it is perhaps surprising that there are only a handful papers on product networks.\(^3\) Dellarocas et al. (2010) study the network of news aggregators. Oestreicher-Singer and Sundararajan (2012) examine how a product’s position in a network affects its demand; we make use of the same networks but instead focus on the network formation. The paper also adds to the studies on firm learning about demand. Hitsch (2006) and Shen and Liu (2014) model how a firm learns from a product’s initial sales after its launch and

\(^{3}\)Some network papers study products that are strongly associated with people, such as user-generated contents. See Mayzlin and Yoganarasimhan (2012), Lu et al. (2013) and Shriver et al. (2013).
exits optimally. Toivanen and Waterson (2005), Shen and Xiao (2014) and Yang (2014) focus on how firms learn from the entry/exit choices of each other in the context of fast food chains. In this paper we look at a different channel of learning, namely from the market outcomes of past products. More broadly, the paper is related to the literature on learning models (Ching et al. (2013)). In terms of the application, the paper looks at the motion picture industry, a popular setting for both marketing and economic research. A wide range of topics have been covered (Eliashberg et al. (2006)). However, there seems not to have been a study modeling the entry decisions where studios green light movies. If so, this paper provides a first attempt, from the perspective of learning.

The rest of the paper is organized as follows. Section 2 describes the data, in particular the construction of the network. Section 3 presents the reduced-form analysis. Section 4 presents the model. Section 5 describes how we estimate the model. Section 6 presents the estimates. Section 7 conducts counter-factual experiments. The last section concludes and discusses future research.

2 Data

Data Sources We mainly use two categories of data. The first include the movie characteristics that are commonly used in studies of motion pictures: title, release date, language, region, genre, MPAA rating, production budget, writers, directors, leading actors, and domestic box-office revenue. Domestic box-office revenue only accounts for a part of a movie’s total revenues. However, it is believed to influence the revenues in subsequent markets, and widely used in measuring the market performance of movies (Eliashberg et al. (2006), Einav (2007)). Because we want to study firm’s go/no go decisions, we also collect data on the production companies and production start date of each movie.

Most of the movie characteristics are collected from the Internet Movie Database (IMDb.com). Additional data on box-office revenues are collected from Boxofficemojo.com, which provides better separation between the revenues from multiple releases if the movie has ever been re-released. In this paper we focus on the box office at the first release. For a small number of movies whose budget sizes are missing on IMDb, we are able to collect them from Wikipedia.com. To account for inflation, we collect data on yearly price level (Consumer Price Index) from the U.S. Bureau of Labor Statistics. To calculate market share from box-office revenues, we collect yearly data on theater ticket price from The-numbers.com, and data on the U.S. population from the Census Bureau.

Production start date is unavailable for roughly one third of the movies. We regress the production period (time elapsed from start date to release date) on the budget size and use the regression to impute the start date. It typically takes slightly more than one year to produce a movie. The estimated relation exhibits an U shape, where the production periods of medium-budget movies

2 Data

Data Sources We mainly use two categories of data. The first include the movie characteristics that are commonly used in studies of motion pictures: title, release date, language, region, genre, MPAA rating, production budget, writers, directors, leading actors, and domestic box-office revenue. Domestic box-office revenue only accounts for a part of a movie’s total revenues. However, it is believed to influence the revenues in subsequent markets, and widely used in measuring the market performance of movies (Eliashberg et al. (2006), Einav (2007)). Because we want to study firm’s go/no go decisions, we also collect data on the production companies and production start date of each movie.

Most of the movie characteristics are collected from the Internet Movie Database (IMDb.com). Additional data on box-office revenues are collected from Boxofficemojo.com, which provides better separation between the revenues from multiple releases if the movie has ever been re-released. In this paper we focus on the box office at the first release. For a small number of movies whose budget sizes are missing on IMDb, we are able to collect them from Wikipedia.com. To account for inflation, we collect data on yearly price level (Consumer Price Index) from the U.S. Bureau of Labor Statistics. To calculate market share from box-office revenues, we collect yearly data on theater ticket price from The-numbers.com, and data on the U.S. population from the Census Bureau.

Production start date is unavailable for roughly one third of the movies. We regress the production period (time elapsed from start date to release date) on the budget size and use the regression to impute the start date. It typically takes slightly more than one year to produce a movie. The estimated relation exhibits an U shape, where the production periods of medium-budget movies
are the shortest.

The second category includes the data on the correlation pattern for the demand across these movies. We look at what two movies tend to be liked by the same consumers. In such case, a high demand for one movie entails a high expectation of demand for the other. Specifically, we make use of the “People Who Liked This Also Liked” feature on IMDb. Through this feature, each movie refers viewers to some other movies. The references are based on an algorithm called item-based collaborative filtering (Desrosiers and Karypis (2011)). In a nutshell, the algorithm calculates a “similarity metric” that measures the correlation between the viewers ratings of the two items. A reference happens when the two items are deemed similar enough. We also make use of “Customers Who Watched This Also Watched” feature on Amazon Instant Video. Slightly different, the similarity metric on Amazon is calculated from user consumption instead of ratings (Linden et al. (2003)).

We define the network by linking two movies whenever there is a reference from one to the other. On IMDb, each movie refers up to 12 other movies; on Amazon, the number is 20. For the analyses shown in the paper, we combine the IMDb and Amazon references. As a robust check, we run the analyses throughout the paper with the network constructed from the Amazon references only, but have not found qualitative differences in the results. The data was collected with a web crawler. Web cookies were disabled to prevent the references from being tailored for the crawling history.

In general, correlation arises when two movies are similar in terms of the characteristics: ideology, story setting, narrative techniques, acting, visual effects, music, etc. Researchers have explored this idea in using consumer purchase data to uncover product positions in a latent characteristic space (Chintagunta (1994), Elrod and Keane (1995), Goettler and Shachar (2001)). It is by this understanding that we use the term “similarity network.” It is possible that correlation arises because of other reasons, for example, complementarity. However, a preliminary examination of the network suggests that this is not the dominant reason, which we turn to below after describing sample selection.

**Sample Selection** We focus on the movies released in the U.S. that started production between 1995 and 2012 (included). The release dates of these movies extend to 2014. We exclude the “micro-budget” movies, which are defined as those with a budget less than 1 million in 2014 dollars. The mechanism behind the production and distribution of the “micro-budget” movies is likely quite different from that of the bigger movies. We have to leave out the movies for which either budget or domestic box-office gross is unavailable. Such movies are typically the ones without significant theatrical release in the U.S. In the end, we have a sample of 3,036 movies.

---

4Some movies are unavailable at Amazon Instant Video. For these movies, only IMDb references are used. However, for our sample, only a small proportion (less than 5%) is not covered.

5The data were collected in March, 2014. The web crawler is based on Scrapy, a open source framework for Python. For more information, visit http://scrapy.org.
It is a good idea to include older movies as the “initial state” for our analyses. This is particularly important for the movies that started closely after 1995 because otherwise they would have no previous similar counterparts and be all mis-regarded as original. Movies that came later in the sample period are less likely to be linked to these earlier movies. We are able to include 1,354 movies from 1975 to 1994 as the initial state. The small sample size is partly due to the fact that fewer movies per year were produced at that time, and partly due to a significant drop in data coverage as we date back before the early 1980s. We have also tried using 1985-1994 as the initial period, but have not found qualitative changes to the results.

The Network  Recall that the links between movies are constructed from references on IMDb and Amazon. Figure 1 provides a visualization of the references with regard to budget size. A dot represents a reference from one movie to another, where the budget of the first movie is given by the horizontal position of the dot, and the budget of the second movie is given by the vertical position. The dots are distributed about evenly on both sides of the 45° line. The plot verifies that the references are based on a symmetric similarity measure, though “Customers Who Watched This Also Watched” makes it sound like that bigger movies are more likely to receive references.

In addition, the plot shows that linked movies tend to have similar budget sizes. This also holds true for the other observed characteristics, as shown in Table 1. For example, among all the possible pairs of movies, 19.4% belong a same genre; the percentage nearly triples when it is among the linked pairs. The last column presents a logit model that predicts whether a pair of movies are linked. All the coefficients are statistically significant. So the movies that are similar in the observed characteristics tend to be linked. Conversely, movies that are dissimilar tend not to be linked. The Pseudo-$R^2$ is .244. Presumably, the unexplained part of the network could be attributed to the unobserved characteristics, such as story setting, narrative techniques, pace, visual effects, sound effects and so on.

We share the concern that the network is constructed from ex post data which studios did not possess when they green-lighted the movies. In respect to this, we do not assume that studios could use the ex post data. Instead, our assumption is that the studios understood the correlation structure represented by the network, and we use the ex post data here to back out that information. In fact, the network that we construct should be transparent enough for an experienced studio executive or movie producer to understand. For example, let us look at Saving Private Ryan: It links to We Were Soldiers and Full Metal Jacket, both action-packed war movies. It also links to Schindler’s List, a WWII movie by the same director. A bit more sophisticatedly, it links to the The Patriot, a movie on American Revolution but by the same writer, and The Shawshank Redemption. However, it does not link to, for example, The English

---

*Both Saving Private Ryan and The Shawshank Redemption belong to the top guy-cry movies selected by Entertainment Weekly, 2005.*
Table 1: Movie Pair Characteristics and Links

<table>
<thead>
<tr>
<th>Feature</th>
<th>All Dyads</th>
<th>Linked Dyads</th>
<th>Logit Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-5.72 (.035)</td>
<td></td>
<td>-5.72 (.035)</td>
</tr>
<tr>
<td>Same Production Company</td>
<td>3.60%</td>
<td>10.6%</td>
<td>0.650 (.048)</td>
</tr>
<tr>
<td>Same Rating</td>
<td>35.3%</td>
<td>59.7%</td>
<td>0.910 (.028)</td>
</tr>
<tr>
<td>Same Genre</td>
<td>19.4%</td>
<td>51.0%</td>
<td>1.26 (.028)</td>
</tr>
<tr>
<td>Same Leading Actor(s)</td>
<td>0.603%</td>
<td>27.8%</td>
<td>3.85 (.034)</td>
</tr>
<tr>
<td>Same Director(s)</td>
<td>0.099%</td>
<td>8.77%</td>
<td>3.53 (.09)</td>
</tr>
<tr>
<td>Same Writer(s)</td>
<td>0.081%</td>
<td>7.26%</td>
<td>2.51 (.09)</td>
</tr>
<tr>
<td>Difference in Release Time</td>
<td>9.76</td>
<td>5.37</td>
<td>-0.113 (.003)</td>
</tr>
<tr>
<td>Difference in Log Budget</td>
<td>1.22</td>
<td>0.762</td>
<td>-0.532 (.020)</td>
</tr>
<tr>
<td>Pseudo-$R^2$</td>
<td>0.244</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>9.63e7</td>
<td>2.91e4</td>
<td>9.63e7</td>
</tr>
</tbody>
</table>

The unit of release time is year. Budget is in 2014 million dollars. In case that there are multiple production companies for one movie, we use the first-listed one. The same applies to genre. The last column is a logistic regression using pair characteristics to predict linkage. Pseudo-$R^2$ equals 1 minus the ratio between residual deviance and null deviance. The entire sample (1975-2012) is included.

![Graph showing the relationship between Referring Movie Budget and Referred Movie Budget.](image)

A dot represents a reference from a movie to another, on either Amazon or IMDb. The horizontal position of the dot gives the budget size of the referring movie; the vertical position gives the budget size of the referred movie. Budget sizes are normalized by CPI to be in 2014 million dollars. The axis scales are nonlinear. The entire sample (1975-2012) is included.

Figure 1: Budgets of the Referred Movies against Referring Movies
*Patient* or *The Reader*, which also use WWII as background but lean toward a more romantic theme.

### 3 Reduced-form Analysis

In this section we present some model-free results. These results are not only interesting in their own right, but also motivate how we write down the structural model in Section 4. At this point, we want to introduce some terminology that will prove useful in the paper. A movie is called a precursor of another movie if the two are linked and the first precedes the second (either with respect to the start dates or the release dates, which will be defined specifically under different contexts). A movie is called an imitator of another movie if the two are linked and the first comes after the second.

First, we are interested in the extent to which the market performance of a movie can be predicted by the market performance of its precursors, in addition to its observed characteristics. Market performance is measured with the ROI, the ratio between domestic box-office revenue and budget. Then we take a look at the studio behaviors and explore what movies are more likely to imitate or be imitated by others.

In Table 2, Column 1 regresses the log ROI on a time trend, genre, MPAA rating, quality of the crew and log budget.\(^7\) The detailed definitions of the covariates are given in the table notes. These covariates are common in the studies of the industry. Notice that there are no significant effects of the “star power,” which is consistent with the finding in Ravid (1999) that stars capture their expected economic rent. Notice that the R\(^2\) of the regression is very low. This is not too surprising: after all, movie success is notoriously difficult to predict. It is worthwhile to point out that though the budget size hardly explains the ROI, it explains substantial variation in the box-office revenue. The R\(^2\) rises to 0.54 if we use the log box-office revenue as the dependent variable, which is comparable to what was found in previous studies.\(^8\)

In Column 2, we add a lag term which equals the average log ROI of the precursors. The coefficient estimate of the lag term is positive and significant, and the R\(^2\) is greatly improved when compared with Column 1. To certain degree, the network controls for the effects of the unobserved characteristics. In particular, notice that coefficient on writer becomes insignificant and much smaller when compared with Column 1. A writer often carries his or her unique style of storytelling from one movie to another, the effects of which seem to have been picked up by the lag term.

In Column 3, we keep the lag term but drop genre, rating and crew as covariates. The decrease

\(^7\)Some readers may be concerned with the fact that budget appears on both sides of the regression. An alternative regression where the dependent variable is replaced by log box-office revenue yields the same coefficient estimates except for that of log budget, which is increased by exactly 1. The same applies to Column 2 and 3.

\(^8\)See, for example, Wallace et al. (1993) and Prag and Casavant (1994). Notice that these studies use smaller samples and include the critical reviews as explanatory variables, which are unavailable before a movie is made.
Table 2: Spatial Regression of Log ROI

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Constant</td>
<td>-0.807**</td>
<td>-0.572**</td>
<td>-0.505**</td>
</tr>
<tr>
<td>Trend</td>
<td>0.0021</td>
<td>0.0089**</td>
<td>0.0107**</td>
</tr>
<tr>
<td>Seasonality</td>
<td>0.130**</td>
<td>0.092**</td>
<td>0.099**</td>
</tr>
<tr>
<td>Log Budget</td>
<td>0.098**</td>
<td>-0.0082</td>
<td>0.0091</td>
</tr>
<tr>
<td>Genre</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Rating</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Crew Actor</td>
<td>-0.198**</td>
<td>-0.106**</td>
<td></td>
</tr>
<tr>
<td>Director</td>
<td>-0.0290</td>
<td>0.0640</td>
<td></td>
</tr>
<tr>
<td>Writer</td>
<td>-0.0644</td>
<td>-0.0186</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0817*</td>
<td>-0.0042</td>
<td></td>
</tr>
<tr>
<td>Average Log ROI of Precursors</td>
<td>0.710**</td>
<td>0.744**</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.058</td>
<td>0.221</td>
<td>0.211</td>
</tr>
<tr>
<td>$N$</td>
<td>2,943</td>
<td>2,943</td>
<td>2,943</td>
</tr>
</tbody>
</table>

** Significant at the 95% level. * Significant at the 80% level. ROI is defined as the ratio between box-office and budget, both of which are normalized by CPI to be in 2014 million dollars. Dependent variable is the log ROI of any movie that started between 1995-2012 and has at least one precursor. A precursor here refers to any similar movie whose release date is earlier than that of the focal movie. Movies in 1975-1994 are used as possible precursors. Trend is the difference in years between the release date and the beginning of 1995. Seasonality uses a dummy for releases in Jun., Jul., Aug. and Dec. Genres are re-categorized into eight “big genres” to reduce the number of parameters. Actor is a dummy for movies with at least one leading actor that had previously taken a leading role in any of the top 5% grossing movies. Director and Writer are defined similarly.

Table 3: Polynomial Fit of Residual Size on Number of Precursors

<table>
<thead>
<tr>
<th></th>
<th>Absolute Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.474 (.0615)</td>
</tr>
<tr>
<td>Number of Precursors</td>
<td>-0.174 (.0259)</td>
</tr>
<tr>
<td>Number of Precursors^2</td>
<td>0.0120 (.0032)</td>
</tr>
<tr>
<td>Number of Precursors^3</td>
<td>-3.43e-4 (1.5e-4)</td>
</tr>
<tr>
<td>Number of Precursors^4</td>
<td>3.40e-6 (2.3e-6)</td>
</tr>
<tr>
<td>Average Effect</td>
<td>-0.0751</td>
</tr>
<tr>
<td>$N$</td>
<td>2,943</td>
</tr>
</tbody>
</table>

Numbers in the parentheses are standard errors. The dependent variable is the absolute value of the residuals from the last column of Table 2. The average effect is the average derivative of the estimated polynomial across the movies in the data.
### Table 4: Regression of the Number of Imitators / Precursors

<table>
<thead>
<tr>
<th></th>
<th>Log # of Precursors</th>
<th>Log # of Imitators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Yearly Dummy</td>
<td>Yes</td>
</tr>
<tr>
<td>Genre</td>
<td>...</td>
<td>Yes</td>
</tr>
<tr>
<td>Rating</td>
<td>Restricted</td>
<td>0.0681*</td>
</tr>
<tr>
<td>Crew</td>
<td>Actor</td>
<td>0.142**</td>
</tr>
<tr>
<td></td>
<td>Director</td>
<td>0.0837**</td>
</tr>
<tr>
<td></td>
<td>Writer</td>
<td>-0.0182</td>
</tr>
<tr>
<td>Log Budget</td>
<td></td>
<td>0.236**</td>
</tr>
<tr>
<td>Log ROI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td>0.379</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>4,390</td>
</tr>
</tbody>
</table>

** Significant at the 95% level. * Significant at the 80% level. See Table 2 for some variable definitions. Here, an imitator of movie $j$ is defined as a movie that is similar to $j$ and started after $j$’s release. A precursor is a movie that is similar to $j$ and released before $j$’s start date. We add 1 to the number of imitators or precursors before taking the log. The entire sample (1975-2012) is included.

The analysis above tells us that the precursors provide a prediction for the ROI, but does not speak to the variance of that prediction. In Table 3 we regress the absolute value of the residuals from the last column of Table 2 on the number of precursors (Using the residuals from Column 2 leads to almost identical results). The estimated average effect is negative, indicating that the prediction variance decreases with the number of precursors. The estimated polynomial actually shows a diminishing decline rate, which is seen in standard Bayesian updating. From the firm’s perspective, this implies that the risks for producing a particular movie decrease with the imitativeness of the movie.

So far our results have been normative. We see that a good deal can be learned from the precursors, however, it is a different matter whether the firms have actually been learning. To take a look into the studio behaviors, Table 4 regresses the log number of precursors and imitators on various movie characteristics and log ROI. Here, a precursor is released before the focal movie’s start date, and an imitator starts after the focal movie’s release. Time dummies are added to control for the fact that the network is truncated outside the sample period. We see that movies with higher ROI are more likely to be imitated, supporting the conventional wisdom that there is firm learning in the movie industry.
Supply of Candidate Movies
Candidates are generated by a stochastic process, and sent to the studios.

Studio Decisions
A Studio is risk averse, manages a portfolio and learns from previous movies.

Consumer Demand
The movie goes into production and realizes its revenue in box office.

Figure 2: Model Overview

A more subtle point in Table 4 is that bigger-budget movies tend to have more precursors as well as imitators. In other words, the network is denser amongst these movies. This is consistent with risk aversion: Budget size multiplies the risks that studios have to face in the box office for a potential movie. To the extent that studios are risk averse, they make the big-budget movies more imitative to keep the risks acceptable. Over time, this makes the network denser among bigger-budget movies. In the structural model, we will allow learning as well as risk aversion on the studio side.

4 Model

In this section we present an empirical model of product entry with learning. An overview of the model is given in Figure 2. First, candidate movies are created continuously over time. Once a candidate lands on the desk of studio executives, they make a go/no go decision based on the market outcomes for previous similar movies. After the movie is made, it goes to theaters where consumers decide how much box-office revenue it will make. We start with the demand side, which is a simple discrete choice model, then introduce the stochastic process that generates the candidate movies, and finally model the studio’s investment decisions.

4.1 Consumer Demand

We model box-office performance. Let vector $x_j$ collect the observed characteristics for movie $j$. Given our data, this can include production budget, genre, MPAA rating, quality of the crew (writers, directors and actors), production company, production start date and release date. A
movie typically stays in theaters for six to eight weeks, with the first two weeks collecting about 60 percent of the lifetime domestic box-office revenue. Let $\xi_j$ capture the average consumer taste for this short period over the characteristics that are not included in $x_j$. Let $\varepsilon_{ij}$ be an idiosyncratic utility term. Consumer $i$’s utility from movie $j$ for this short period after $j$’s release is

$$ u_{ij} = U(x_j; \beta) + \xi_j + \varepsilon_{ij}. $$

We explicitly model the correlation across the $\xi$’s for different movies. The correlation structure is captured with the network, which we treat with more details when presenting the supply side. Generally, we can think that the correlation arises when the two movies are similar in characteristics. To the extent that $\xi_j$ measures the excellence of the movie in the eye of consumers, we refer to it as the “latent quality.”

In the reduced-form analysis that, we have seen that once the precursors are accounted for, characteristics such as genre, MPAA rating and quality of the crew add very little prediction power. We will see very similar results for the estimates of the demand model, presented later in Section 6. For this reason and for tractability, these characteristics are not included in $x_j$ for the benchmark model, in particular the supply side. In addition, not all the elements of $x_j$ need to enter $U(\cdot)$. For example, it is probably far-fetched to argue that the production company or production start date would enter consumer utility.

To complete the demand model, suppose that individual $i$ chooses between going to a movie theater to watch $j$ and an “outside option,” for which we specify the utility as

$$ u_{i0} = \varepsilon_{i0}. $$

Then, assuming type-I extreme value distribution for the idiosyncratic errors $\varepsilon_{ij}$ and $\varepsilon_{i0}$, we have the market share of $j$ given by $1/(1 + e^{-U(x_j; \beta) - \xi_j})$. To convert market share into revenue, we multiply it by the market size and average ticket price at theaters. The market size is taken as the population of “moviegoers” who go to cinema at least once a year, about two-third of the population.\(^9\) Let $m_t$ be the multiplier at time $t$, and $r_j$ the release date of $j$. The box-office revenue for $j$ is given by\(^10\)

$$ \pi_j = m_{r_j} / (1 + e^{-U(x_j; \beta) - \xi_j}). \quad (1) $$

Note that there is a one-to-one relation between the box-office revenue $\pi_j$ and latent quality $\xi_j$.

\(^9\)See *Theatrical Market Statistics*, MPAA. We treat both the market size and ticket price as exogenous time series. It is an known fact (as well as a puzzle) that theatrical ticket price hardly varies across seasons and movies. See Orbach and Einav (2007) for more discussions.

\(^10\)The model abstracts away from several factors that can affect demand, including the marketing expenditure, number of screens and concurrent releases. These factors are determined after the movie is made, and are endogenous outcomes of budget size, movie quality and competition at theaters. See Hennig-Thurau et al. (2006) for the relative importance of marketing vs. movie quality. They find that overall, quality is more important. See Elberse and Eliashberg (2003) for exhibition dynamics, Ainslie et al. (2005) for market share competition, and Einav (2007, 2010) for release competition and timing.
So after a movie is exhibited in theaters, the box-office performance reveals the latent quality of the movie.

### 4.2 Candidate Products

Instead of pre-fixing a product space where firms can choose from, we use a stochastic arrival process to generate the candidates for the firms. The reasons for this are twofold: From the substantive point of view, this captures the finite supply of potential movies where not all conceivable movies are available at all times. From a technical point of view, this reduces the dimension of the firm’s problem, permitting a tractable model.

Treating time as continuous, we let candidate movies arrive at a Poisson rate $\eta_f$ for firm $f$.\[^{11}\]

Suppose that a candidate movie $j$ arrives at time $t$ for firm $f$. If it ever gets produced, we record its arrival time by $a_j$ and its production period is $[a_j, r_j]$. The observed characteristics $x_j$, latent quality $\xi_j$, as well as the location of $j$ in the similarity network, are drawn from a state-dependent distribution. The state is denoted as $S_t$ and is the collection of the observed characteristics, latent qualities and network for all the existing movies at $t$ (released or in-production).

We first draw the candidate’s location in the network, namely what existing movies are similar to $j$. Then conditional on the location, we draw the characteristics $x_j$ and $\xi_j$. In this sense, each candidate is either completely novel or a “creative combination” of some existing products.\[^{12}\]

In principle, one can also first draw characteristics and then determine the location by its distances to the other products in the characteristic space. But this does not seem empirically implementable because many characteristics are unobserved in the data.

Before continuing to the specification of the arrival distribution, we want to point out that the arrival process is a latent structure. We have experimented with many variations of the process and our choice has been guided by both economic intuition and patterns in the data. The extent to which the model is capable of reproducing the data is partially assessed in Section 6.2.

**Network Location** The set of existing movies at time $t$, $\{k : a_k < t\}$, includes those that are either released or still in production. The similarity amongst these movies is described by the network at the time. The location of the candidate $j$ in the network is described by which existing movies are linked to $j$. Formally, we use $y_j$ to denote the location for $j$. It is a vector of the length of the number of existing movies, where $y_{k,j} = 1$ indicates that a link is formed between $j$ and the existing movie $k$, and $y_{k,j} = 0$ otherwise. In the language of evolving network models, the candidate “attaches” to the existing network. We let the attachment probability

\[^{11}\]The model allows a different arrival rate for each firm, which introduces quite a lot parameters. For estimation, we use a single arrival rate and assign each arrival to a firm with the probability proportional to its share of movies in the data. For each movie, the first listed production company is counted as the firm for that movie.

\[^{12}\]Combinatorial creativity has been proposed and studied in multiple areas, including psychology, economics and science in general, see Mednick (1962), Weitzman (1998) and Uzzi (2013).
follow a logit model:

$$\Pr(y_{k,j} = 1|S_t) = \frac{1}{1 + e^{-F(x_k,t;\gamma)}}, \quad \forall k.$$ (2)

We specify $F(x_k,t;\gamma) = \gamma_0 + \gamma_1[k \in f] - \gamma_2(t - a_k)$. The first term is a constant. The second term is an indicator dummy that gives potentially higher probability to attachment towards the movies produced by firm $f$. It captures the possibility that a firm has a favorable position to imitate its own movies. For example, it may have developed exclusive relations with the directors in its past movies. The third term discounts movies by their arrival dates. So any movie gradually becomes obsolete and unlikely to be imitated anymore. These two specifications are consistent with the properties of the observed network that we showed with Table 1.

If the attachments to the existing movies are independent of each other, it is simple to draw the location for $j$. Unfortunately, this should not be the case. If two existing movies are similar to each other, it should be unlikely for $j$ to link to only one of them. In other words, it is more likely for us to see a complete triangle among the candidate and the two movies rather than an incomplete one missing an edge. The prevalence of triangles in networks is often called clustering. In social networks, clustering refers to the property that “my friends are friends themselves.” There is substantial clustering for the network in our data. The average clustering coefficient$^{13}$ is 0.22. As a comparison, randomly assigning the same number of links to the same number of nodes yields a clustering coefficient typically less than 0.01.

We introduce correlation in the attachments through a two-stage process. A similar process was proposed in physics (Holme and Kim (2002)) as a simple but flexible way to generate clustering for general networks. The same idea has been used to fit the clustering in social networks (Jackson and Rogers (2005)). Specifically, in the first stage, $j$ forms link with each

---

$^{13}$The clustering coefficient of a node equals the number of triangles that it belongs to divided by the number of triangles that it would belong to if all of its neighbors were linked with each other. See Watts and Strogatz (1998).
existing movie independently. In the second stage, for each \( k \) linked in the first stage, \( j \) further forms links with each neighbor of \( k \) with probability \( \omega \). An example of the two-stage process is illustrated in Figure 3. In the Appendix we show how to calculate the 1st-stage probabilities from (2).

**Observed Characteristics** We have determined the network location for \( j \). Now we specify a distribution \( \Pr(x_j|y_j, S_t) \) from which the observed characteristics of \( j \) is drawn. Notice, in particular, that the distribution depends on \( y_j \). This allows \( x_j \) to be correlated with the characteristics of its precursors. For example, a candidate that is similar to a group of big-budget movies should likely be a big-budget. Had we included genre in \( x_j \) on the supply side, a candidate whose precursors are all sci-fi movies should probably be a sci-fi as well.

Specifically, the budget \( b_j \) is drawn from a truncated normal distribution. The variance-to-mean ratio, which we denote by \( \chi \), is to be estimated as a parameter. If the set of the precursors for \( j \) is nonempty, the mean of the truncated normal is set equal to the average budget of the precursors. Otherwise the mean equals \( \mu \), which is another parameter to be estimated.\(^{14}\) The release time is determined by \( r_j = a_j + \Upsilon(b_j) \), where \( \Upsilon \) is a function that relates production time with budget size, and is nonparametrically estimated “off-line” with the data on production start date. This is a simplification and we abstract away from the factors that influence the release date after the green-light decision.

**Latent Quality** Now we specify a distribution \( \Pr(\xi_j|x_j, y_j, S_t) \) from which the latent quality of \( j \) is drawn. Here we want to consider two factors. First, we want \( \xi_j \) to be correlated with \( \xi_k \) if \( k \) is a precursor of \( j \). Second, recall that the latent quality measures the consumer tastes at the time of the movie’s release. To the extent that consumer tastes can be changing over time, we want to allow the difference in the release time, \( |r_j - r_k| \), to dilute the correlation between \( \xi_j \) and \( \xi_k \). It seems appropriate to make the following specification:

\[
\Pr(\xi_j|x_j, y_j, S_t) = \mathcal{N}\left( \frac{\lambda \sum_{y_{k,j}=1} \varphi|r_k-r_j| \xi_k}{1 + \lambda \sum_{y_{k,j}=1} \varphi|r_k-r_j|}, \frac{\sigma^2}{1 + \lambda \sum_{y_{k,j}=1} \varphi|r_k-r_j|} \right),
\]

Note that (3) resembles the Bayesian updating formula under normality, where the latent quality of a precursor, \( \xi_k \), is treated as a signal for \( \xi_j \). The weight for each signal is calibrated by \( \lambda \), which can be thought of as a measure of correlation. For the extreme case \( \lambda = 0 \), the \( \xi \)'s are independent. The weight is discounted by the difference in release time. For the extreme case \( \varphi = 0 \), consumer tastes change so rapidly that two similar movies that are released at different times, even if close, will have completely uncorrelated market outcomes.

\(^{14}\)In the data, budget distributes around the mean budget of precursors in a truncated normal shape, and the dispersion hardly shrinks with the number of precursors. We use \([1, 350]\) as the truncation interval, as the biggest budget observed in the data is $343m. Results are not sensitive to the choice of the interval upper bound.
4.3 Go/No Go Decision

When it is the time to green-light a potential movie, studio sees the story and screenplay, and in most cases has a reliable estimate of the budget and release date. The producer often has secured some of the crew and is aware of who else she or he needs to recruit. However, much uncertainty remains on how this particular movie will be received in the box office.\footnote{Here is a description of the “green-light” process by a senior studio executive: “We bring together all studio department heads. [The production costs] is our most reliable estimate, and that thus forms the basis for our launch decision.... In the end; ... Someone in the meeting has to put his or her reputation on the line and say ‘yes’ - regardless of whether the numbers add up” Eliashberg et al. (2006)} Our corresponding assumption is that firms do not observe the latent quality for either the candidate or any movie that is still in production. However, we do assume that the firms know the demand correlation across movies, namely the network, which they use to form belief on the unknown latent qualities.

To formally model the go/no go decisions, we start with the information set for the firms, denoted as $\mathcal{F}_t$. The set includes the observed characteristics and the network of the existing movies as well as the candidate, if there is one arriving at $t$. $\mathcal{F}_t$ also includes the box-office revenue $\pi_k$, or equivalently, the latent quality $\xi_k$ of each movie $k$ that has been released. Notice the important difference between $\mathcal{S}_t$ and $\mathcal{F}_t$ that the later does not include the latent qualities of the movies that are still in production.

Given the information set, we can work out the belief for the firms. For the simplest case where we are looking at a single candidate $j$ whose precursors are all released, the belief $\Pr(\xi_j|\mathcal{F}_t)$ is simply given by (3). However, it is possible that some precursor has not been released. In such case, one can learn indirectly from the released neighbors of that precursor, whether they arrived before or after that precursor. In addition, if $j$ and the precursor belong to the same firm, then the correlation between the market outcomes for these two movies will amplify the risks associated with producing $j$. So in general, it is important for us to look at the joint belief: $\Pr(\{\xi_k : a_k \leq t, r_k \geq t\}|\mathcal{F}_t)$. A nice feature of our model is that there is a closed-form expression for this joint density, which we derive in the Appendix.

Given the belief, now we can model the firm’s investment decisions. Given $\mathcal{F}_t$, we denote by $P_{f,t} := \{k \in f : a_k < t, r_k \geq t\}$ the set of $f$’s in-production movies, which can be thought of as $f$’s portfolio. We ask the firm if it is desirable to add $j$ into the portfolio. The present value of $P_{f,t}$ is

$$\Pi(P_{f,t}) = \sum_{k \in P_{f,t}} \delta^{r_k - t} \pi_k,$$

where $\delta$ is a discounting factor and $\pi_k$ is given by equation (1). If $j$ is accepted, the present value of the new portfolio becomes

$$\Pi(P_{f,t} \cup \{j\}) = \delta^{r_j - t} \pi_j + \Pi(P_{f,t}).$$

At time $t$, these present values are uncertain to the firm because the $\pi$’s depend on the $\xi$’s, which
the firms does not know. Let $b_j$ be the production budget of $j$, and $\zeta_j$ an independent decision error that captures the factors unobserved to us but known to the firm. Let $V$ be a concave function that represents the valuation by a risk-averse firm. Candidate $j$ is accepted iff

$$E \left[ V \left( \Pi(P_{f,t} \cup \{j\}) - b_j - \zeta_j \right) \bigg| \mathcal{F}_t \right] > E \left[ V(\Pi(P_{f,t})) \bigg| \mathcal{F}_t \right],$$

(4)

where the expectation is taken over the $\xi$’s. We specify $V$ to bear a constant coefficient of absolute risk aversion (CARA), denoted as $\alpha$. We specify $\zeta_j = (e^{z_j} - 1)b_j$ where $z_j$ is distributed $\mathcal{N}(0, \rho^2)$. The firm discards the candidate if condition (4) does not hold.

An important feature of this formulation of firm’s problem is that it takes into account the revenue correlation across the movies in $P_{f,t}$. To the extent that the firm is risk averse, it would like to “diversify” its portfolio and avoid investments in many similar movies at once. However, the formulation treats the firm myopic, not taking into account how a decision today will affect future arrivals and decisions. To solve for a full model of forward-looking decisions with a network structure is beyond this paper. We leave it for future research.

We can readily define a risk-free equivalence of the revenue $\pi_j$, denoted by $\pi_j$, through the equation

$$E \left[ V(\Pi(P_{f,t} + \pi_j)) \bigg| \mathcal{F}_t \right] = E \left[ V(\Pi(P_{f,t} \cup \{j\})) \bigg| \mathcal{F}_t \right].$$

When $P_{f,t}$ is empty, it reduces to the more familiar equation: $E(V(\pi_j)\big|\mathcal{F}_t) = V(\pi_j)$. The definition allows us to state condition (4) alternatively as $\pi_j/b_j - \zeta_j > 0$. We may also define $\pi_j/b_j$ as the risk-adjusted ROI for movie $j$. From the perspective of the econometrician, the probability that a candidate will be accepted is monotone in the risk-adjusted ROI.

An effective assumption here is that a movie with $\pi_j/b_j = 1$ is accepted with .5 probability. This is a normalization. The journey of a movie often goes beyond production and domestic box office, incurring additional expenditures on advertising and exhibition, while earning additional revenues from home video sales and international markets. So it is entirely possible that the studio is not at all indifferent between accepting and rejecting such a movie. However, with the arrival process being latent, the model is observationally equivalent if we halve the acceptance probability for every candidate and double the arrival rate at the same time. Effectively, we can only identify the relative acceptance probabilities for different movies. In the Appendix, we use a simplified version of the model to further explain why the normalization is innocuous and necessary.
5 Model Estimation

5.1 Demand

Although the products in the market are consequences of firm selection, most applications estimate demand by assuming that the set of products is exogenous and focus on other sources of endogeneity (e.g., price). Even in studies of market or product entry, it is standard to retain exogeneity in the unobserved component $\xi$ by arguing that firms have no knowledge of it before entry (see, for example, Aguirregabiria and Ho (2011) and Eizenberg (2014)). Because our model relaxes this assumption, it requires an extension of the standard estimation technique.\footnote{An important difference here from the standard spatial econometrics is that the network is not exogenous. For general treatment of spatial econometrics, see Bradlow (2005) and LeSage (2008).}

To be more specific, notice that the following regression equation can be directly obtained from the box-office equation (1).

$$\log (\pi_j) - \log (m_{r_j} - \pi_j) = U(x_j, \beta) + \xi_j. \quad (5)$$

There are two issues about this regression. First, the residuals $\xi$ are correlated, and the correlation structure is of interest to us. Second, due to endogenous entry, the standard moment condition $E(\xi_j|x_j) = 0$ does not hold here in general. For example, $\xi_j$ can be positively correlated with the budget $b_j$ in $x_j$, because a bigger budget implies larger risks which typically need to be compensated by a higher belief on $\xi_j$ for entry.

We estimate this equation by controlling for what firms can learn about $\xi_j$ at the time of entry. In our model, the firm’s information at time $a_j$ is a subset of $(x_j, y_j, S_{a_j})$, which determines the arrival distribution of $\xi_j$ through (3). Let $\nu_j \equiv \xi_j - E(\xi_j|x_j, y_j, S_{a_j}; \beta, \lambda, \varphi)$, which is the difference between the realized latent quality and the mean of its arrival distribution. We use $\nu$ instead of $\xi$ to construct moment conditions. Identification of parameter $\sigma$ requires us to further match the dispersion of the arrival distribution, so we define a second difference: $\iota_j \equiv \nu_j^2 - E(\nu_j^2|x_j, y_j, S_{a_j}; \beta, \lambda, \varphi, \sigma)$. Our demand-side estimation are then based on the following mean-independence moments:

$$E\left( (\nu_j, \iota_j) | x_j, y_j, S_{a_j} \right) = 0.$$

This leaves us with many instruments to choose from the conditioning set to interact with $\nu_j$ and $\iota_j$.\footnote{For reasons why we do not use many moment conditions, see Andersen and Sørensen (1996) and more recently Han and Phillips (2006).} To identify $\beta$, we interact $\nu_j$ with $x_j$. To identify parameter $\lambda$ and $\varphi$, we interact $\nu_j$ with the average latent quality of the precursors for $j$, and the precursors that were released several years apart from $r_j$. To identify $\sigma$, we interact $\iota_j$ with a constant term and the number of the precursors for $j$. To the extend that the $\xi$’s are unobserved, they need to be computed through equation (5) as a function of the unknown parameter $\beta$. This means that the estimation requires
a numerical search jointly over \((\beta, \lambda, \varphi, \sigma)\). We save lots of computational time by using the OLS estimates of (5) as the initial guess for \(\beta\).

The sample moments average across the movies that started production in 1995 and after. The movies in 1975-1994 are counted as possible precursors of these movies. Not conditioning on this initial sample should not affect the asymptotics of the estimator as the sample period extends, but introduces a source of finite-sample bias.

### 5.2 Supply

The estimation procedure for the supply side is relatively straightforward. We match the properties that the model predicts for the *produced* movies with those observed in the data. Specifically, index the movies in data by arrival date so that \(j\) is the first movie that arrives after \(j - 1\). The full history up to time \(a_j\) can be summarized as \((x_j, y_j, \xi_j, S_{a_j})\). Let \(H\) be a function this history. For notation, we write \(H(x_j, y_j, \xi_j, S_{a_j})\) as \(H_j\). The specification of \(H\) depends by the moments that we want to match to identify the parameters. We give the specification of \(H\) below after discussing the identification.

Collect the supply-side parameters in \(\Theta\). For any value of \(\Theta\), given the history at time \(a_j - 1\), our model makes a prediction of \(H_j\), where the error of prediction given by

\[
h(x_j, y_j, \xi_j, S_{a_j}; \Theta) \equiv H_j - \mathbb{E}(H_j|x_{j-1}, y_{j-1}, \xi_{j-1}, S_{a_{j-1}}; \Theta).
\]

The conditional expectation does not have closed forms, but can be evaluated through simulations. This evaluation step is computationally intensive and required us to make use of a computer cluster.\(^{18}\) Our supply-side estimation relies on the following moment conditions:

\[
\mathbb{E}\left(h(x_j, y_j, \xi_j, S_{a_j}; \Theta)\right) = 0.
\]

The estimate for \(\Theta\) is obtained following the standard procedure of the Generalized Method of Moments. It searches for the parameter values that minimizes a norm of the sample counterpart of the moment conditions: \(\frac{1}{n-k+1} \sum_{j=k}^{n} h(\cdot)\) where \(k\) is the first movie produced in 1995. Again, movies in 1975-1994 are counted as possible precursors but not included in sample moments.

In our model, the arrival process is latent, and the set of produced movies is a joint outcome of the arrival process and the studio selections. Consequently, identifications for some of the supply-side parameters are not straightforward. Here we provide some intuition as to how these parameters are identified. In the Appendix, identification is shown for a simplified version of the model; for the full model, we show through Monte Carlo experiments that the parameter values can be recovered from simulated data.

\(^{18}\)It is not necessary to use a very large number of simulations to evaluate one conditional expectation, as the simulation errors will be averaged across the movies. We chose to use 100 simulations. Nevertheless, one evaluation of the objective function takes around 10 mins on a quad-core desktop.
For the arrival process, the second-stage attachment probability ($\omega$) is identified by the extent of clustering in the observed network. Although the clustering coefficient can be affected by other parts of the model, ceteris paribus, it should always increase with $\omega$. The identification of the arrival rate ($\eta$) relies on the normalization that a movie with $\pi_j = b_j$ is accepted with $.5$ probability. As already pointed out, the model is observationally equivalent if we halve the acceptance probability for each candidate and double the arrival rate at the same time. In this regard, we should not interpret the estimated $\eta$ literally as the number of the proposals that are sent to studios over a year.

For the selection model, the coefficient of risk aversion ($\alpha$) can be identified by the difference in the imitativeness between big-budget and small-budget movies. To the extend that the firms are risk averse, they make bigger-budget more imitative to keep the risks acceptable. This is also in line with our reduced-form findings. Alternatively, risk aversion can be identified simply by the mean of the observed budget distribution, given the model assumption that the budget of a candidate centers at the mean budget of the precursors. This is because the acceptance probability for big-budget movies tends to decrease with risk aversion. The other parameter is the standard deviation of the decision error ($\rho$). It can be identified by the average risk-adjusted ROI of the movies in the data. This is because the firms become more selective when $\rho$ decreases.

The specification of $H$ complies with our identification argument. Specifically, we include in $H_j$: the time elapsed since last movie production ($a_j - a_{j-1}$), an indicator whether $j$ has any precursors, the log number of precursors, number of the triangles created in the attachment of $j$ divided by the number of precursors squared, proportion of the precursors that were produced by the same firm as that of $j$, average age of the precursors at time $a_j$, log of budget $b_j$, distance between $b_j$ and the empirical mean of budget ($$48m$), distance between $b_j$ and the average budget of the precursors, and finally the log of the risk-adjusted ROI of $j$.

### 6 Estimation Results

#### 6.1 Parameter Values

**Demand side** Table 5 displays the estimates for the demand-side parameters. Specifically, Column I displays the estimates from an OLS regression of the revenue equation (5); Column II displays the GMM estimates with all the covariates; Column III displays the GMM estimates excluding genre, MPAA rating and quality of the crew as covariates.

First notice that the estimates do not differ too much across the three configurations, so some common observations can be made. There is a small but statistically significant downward trend, which may be attributed to the growth of the home video market as an alternative to movie theaters. The demand tends to be higher in the summer and at the end of a year, which is consistent with the seasonality pattern found in Einav (2007). As expected, a “restricted” MPAA
Table 5: Model Parameter Estimates, Demand-Side

<table>
<thead>
<tr>
<th>Parameters</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Constant</td>
<td>-7.14 (.14)</td>
<td>-7.29 (.18)</td>
<td>-7.32 (.17)</td>
</tr>
<tr>
<td>Trend</td>
<td>-0.0202 (.004)</td>
<td>-0.0188 (.008)</td>
<td>-0.0153 (.008)</td>
</tr>
<tr>
<td>Seasonality</td>
<td>0.151 (.05)</td>
<td>0.110 (.04)</td>
<td>0.119 (.04)</td>
</tr>
<tr>
<td>Budget</td>
<td>Yes (Fig. 4)</td>
<td>Yes (Fig. 4)</td>
<td>Yes</td>
</tr>
<tr>
<td>Rating</td>
<td>Restricted</td>
<td>-0.198 (.05)</td>
<td>-0.207 (.06)</td>
</tr>
<tr>
<td>Genre</td>
<td>Drama</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Comedy</td>
<td>0.418 (.07)</td>
<td>0.238 (.08)</td>
<td></td>
</tr>
<tr>
<td>Action/War</td>
<td>0.197 (.07)</td>
<td>0.183 (.08)</td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td>0.385 (.09)</td>
<td>0.435 (.15)</td>
<td></td>
</tr>
<tr>
<td>Sci-Fi/Advent.</td>
<td>0.358 (.1)</td>
<td>0.314 (.09)</td>
<td></td>
</tr>
<tr>
<td>Horror</td>
<td>1.11 (.1)</td>
<td>0.962 (.2)</td>
<td></td>
</tr>
<tr>
<td>History/Bio.</td>
<td>0.107 (.1)</td>
<td>0.109 (.1)</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>-0.426 (.4)</td>
<td>-0.680 (.3)</td>
<td></td>
</tr>
<tr>
<td>Crew Power</td>
<td>Actor</td>
<td>-0.0264 (.05)</td>
<td>0.00823 (.05)</td>
</tr>
<tr>
<td></td>
<td>Director</td>
<td>-0.0677 (.06)</td>
<td>-0.0301 (.05)</td>
</tr>
<tr>
<td></td>
<td>Writer</td>
<td>0.121 (.06)</td>
<td>0.0493 (.05)</td>
</tr>
<tr>
<td>Similarity ($\lambda$)</td>
<td></td>
<td>0.466 (.06)</td>
<td>0.529 (.06)</td>
</tr>
<tr>
<td>Disc. Factor ($\varphi$)</td>
<td></td>
<td>0.929 (.02)</td>
<td>0.944 (.02)</td>
</tr>
<tr>
<td>Std. Dev. ($\sigma$)</td>
<td></td>
<td>1.82 (.05)</td>
<td>1.87 (.05)</td>
</tr>
<tr>
<td>$R^2$ (Share)</td>
<td>0.557</td>
<td>0.657</td>
<td>0.653</td>
</tr>
<tr>
<td>$R^2$ (ROI)</td>
<td>0.0716</td>
<td>0.282</td>
<td>0.272</td>
</tr>
</tbody>
</table>

Column I displays the OLS estimates of equation (5). Column II and III display the GMM estimates. See Table 2 for definitions of some of the variables. The utility for budget is estimated as a piecewise linear function; see Figure 4 for the estimates. Discounting factor $\varphi$ is yearly. The numbers in parentheses are standard errors. $R^2$(share) measures the fit for the log market shares. $R^2$(ROI) measures the fit for the log ROIs.

![Figure 4: Estimated Effects ($\beta$) for Budget Size](image)

The solid curve is the GMM estimate of the utility for budget (Column II in Table 5). The dashed curve is the OLS estimate (Column I in Table 5). The utility is specified as a piecewise linear function. The shape displays diminishing marginal utility. However, the diminishing rate is slower than that of a logarithm specification.
rating reduces demand. Interestingly, horror movies are the best bet for studios to make profits.\textsuperscript{19} The effect of star power is insignificant, which is consistent with our reduced-form analysis and the finding in Ravid (1999) that stars capture their economic rent. Finally, the effects of budget are estimated as a piecewise linear function and plotted in Figure 4. The shapes of the function exhibit diminishing marginal utility.

The difference between the GMM and OLS estimates for the effects of budget (Figure 4) can be explained by studio selection. For example, for a big-budget movie to be produced, a high belief on $\xi$ is typically required to compensate the associated large risks. This introduces a positive correlation between $b_j$ and $\xi_j$, making the OLS estimates biased towards larger effects of budget. The estimated effects of genre, rating and quality of the crew tend to be smaller with the GMM (Table 5). This is because these effects are taken into account, to certain extent, by the network.

As to the explanatory power, by using the network, Column II improves the $R^2$ for market share from the .56 in Column I to about .66. Notice that budget size is a major explainer for market share and so contributes significantly to the $R^2$. In terms of explaining the ROI, we see that the model in Column I performs poorly, whereas Column II provides a considerable improvement. This is in line with our reduced-form results (Table 2). As we move from Column II to III, both the $R^2$'s remain almost identical, and $\lambda$ picks up the effects of the dropped covariates. This is again in line with our reduced-form analysis, indicating that the network incorporates the similarity in these covariates. The result justifies us to drop these covariates on the supply side, making the model and estimation much more tractable.

The estimate of $\sigma$ implies an enormous level of uncertainty. To see the magnitude, recall that $\sigma$ is the standard deviation of the latent quality for a candidate without any precursors; one standard deviation equals about the effect of raising the budget of a $10m movie to $70m, or the budget of a $100m movie to over $300m. Under the estimates of $\lambda$ and $\varphi$, our model implies that for the movies in the data, learning on average reduces the variance in $\xi$ at the time of go/no go decision by about 60%\textsuperscript{20}. The estimate of $\varphi$ indicates a fairly rapid change of consumer tastes over time. In updating on the latent quality for a candidate, a 10-year old precursor counts only slightly more than half as much as a concurrent precursor.

Supply side Table 6 displays the supply-side estimates. First shown are the parameters for the attachment process. We see a sizable 2nd-stage attachment probability, which is consistent with the degree of clustering in the observed network. The estimate for $\gamma_1$ indicates that a production company has certain monopoly power in imitating its own movies. A studio may become specialized in certain types of movies, and often develops exclusive relation with the crew from its past productions.

\textsuperscript{19}For a stimulating discussion on this, see “Let’s Get Scared: Why Horror Movies Are Immune to the Digital Onslaught.” September 16, 2013, Yahoo Movies.

\textsuperscript{20}More precisely, for each movie, we compare $\text{Var}(\xi_j | \mathcal{F}_{\tau_j})$ with $\sigma^2$, which is the variance that the firm would perceive had it ignored the correlation structure in $\xi$. \hfill 22
Table 6: Supply-Side Parameter Estimates

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attachment</td>
<td>2nd-Stage Probability ($\omega$)</td>
</tr>
<tr>
<td>Interception ($\gamma_0$)</td>
<td>-4.836 (.02)</td>
</tr>
<tr>
<td>Own Movies ($\gamma_1$)</td>
<td>1.673 (.06)</td>
</tr>
<tr>
<td>Time Difference in Years ($\gamma_2$)</td>
<td>-0.214 (.005)</td>
</tr>
<tr>
<td>Obs. Characteristics</td>
<td>Budget Mean without Precursors ($\mu$)</td>
</tr>
<tr>
<td></td>
<td>Variance-to-Mean Ratio ($\chi$)</td>
</tr>
<tr>
<td>Coefficient of Risk Aversion ($\alpha$)</td>
<td>0.0312 (.0051)</td>
</tr>
<tr>
<td>Standard Deviation of Decision Shock ($\rho$)</td>
<td>0.601 (.03)</td>
</tr>
<tr>
<td>Yearly Arrival Rate ($\gamma$)</td>
<td>772 (45)</td>
</tr>
</tbody>
</table>

Budgets are expressed in 2014 million dollars. A single arrival rate is estimated where any arrival is assigned to one firm according to its share of movies in the data. The firm discounting factor $\delta$ is set at .975. Numbers in the parentheses are standard errors computed by parametric bootstrapping (see Appendix).

The plots show the risk-adjusted ROI, $\pi_j/b_j$, of a hypothetical movie $j$ as a function of the budget size and variance in the latent quality $\xi_j$. The latent qualities of the precursors for $j$ are set equal to the average latent quality of the movies with budget close to $b_j$ in the data. Budget is again expressed in 2014 million dollars. The plot on the left uses model estimates, while the plot on the right takes $\alpha \to 0$ so firms are risk-neutral. ROI only takes domestic box office and production budget into account.

Figure 5: Risk-adjusted ROI as Function of Budget Size and Var($\mu$)
The coefficient of absolute risk aversion, $\alpha$, is estimated to be both statistically and economically significant. This is demonstrated in Figure 5. On the left side, the graph plots the risk-adjusted ROI as a mapping of the budget size and variance in the $\xi$. We see that for big-budget movies, risk-adjusted ROI decreases with the variance in $\xi$. This should be expected because a higher variance implies a higher level of risks. However, what may come as a surprise is that this relation is reversed for small-budget movies. This is because there is a tight lower bound for how much a small-budget movie can lose: the maximum that a movie with $5$ million budget can lose is $5$ million. In this case, a larger variance expands the right tail of the box-office revenue distribution but not as much for the left tail. In other words, for a small-budget, novelty increases the chance of becoming a big hit. This mapping for risk-adjusted ROI is a driving force behind some of the predictions later in the counter-factual analysis.\(^\text{21}\)

As a comparison, the graph on the right side in Figure 5 plots the risk-neutral case where $\alpha \to 0$. The graph depicts a very different mapping, indicating that risk aversion does play a significant role. In particular, for almost all levels of budget size, firms prefer novel candidates, taking advantage of the lower bound of the box-office revenue distribution.

Recall that we allow for a decision shock on the firm side, capturing factors not observed by us but known to the firms. The larger is the size of the shock, $\rho$, the less predictive is our model on the firm decisions. In the extreme case where $\rho \to +\infty$, all types of arrivals are accepted with equal probability. Under the estimate of $\rho$, the acceptance probability ranges from about 0 to .55 for the range of the risk-adjusted ROIs shown in the left graph of Figure 5. So our model captures a good deal of the firm decisions. Finally, the estimate of the arrival rate implies that around three quarters of the arrivals are rejected. Because the identification of the arrival rate relies on a normalization, the estimate should not be interpreted literally as the number of candidates that are sent to the studios each year.\(^\text{22}\)

6.2 Model Fit

To make an assessment on model fit, we simulate the model from 1995 all the way to 2012, conditional on the data from 1975 to 1994. In Figure 6, we compare the simulated data with the real data. Given the important trade-off between budget size and uncertainty in the firm’s decision (see Figure 5), here we look at: (i) the scatter plot of the number of precursors against the budget, (ii) the distribution of the number of precursors, (iii) the budget distributions, and (iv) the distribution of ROI.

Considering the parsimony of our supply model, we find the fit satisfactory. Because the produc-

\(^{21}\)Given this mapping, one may ask why not split the money for a novel big-budget into many small-budget and novel movies? The immediate answer is that movie supply is not infinite. Once there are many small-budget movies produced, it becomes difficult to come up with another original small-budget. Goettler and Leslie (2005) asked a similar question and offered a few alternative explanations.

\(^{22}\)For readers interested in the transaction of movie scripts, see Luo (2014). However, rejected scripts are not included in her data and no estimate of the rejection rate is provided.
The model is simulated for once from 1995-2012 conditional on the data up until 1994. The column on the left plots the simulated data, while the column on the right plots the real data. Each row plots, respectively: (i) the log number of precursors against the log budget, (ii) the distribution of the log number of precursors, (iii) the distribution of the log budget distribution, (iv) the distribution of log ROI. A precursor for $j$ is any $k$ that is linked to $j$ and satisfies $a_k < a_j$. We add 1 to the number of precursors before taking its log.

Figure 6: Model Fit
tion and release strategies can be different across movies with diverse sizes, production companies and release years, it is difficult for the model to capture all the patterns in the data. For example, the model seems to under-produce the very big-budget movies. This could be caused by risk aversion heterogeneity across firms, which our model fails to capture: blockbusters are often produced by major studios that are financially more capable than independent production companies. The model also seems to produce a smaller left tail for the log ROI. This could be caused by the normality assumption on $\xi$. The fatter left tail in the data suggests that it may be better to use a distribution that allows some degree of negative skewness. Enriching the model for a better fit with the data is left for future research.

7 Counter-factual Experiments

In this section we use several counter-factual experiments to provide further insights on how learning affects product entry. First, we quantify the importance of learning in our model by investigating what happens if firms stop learning. We also try to quantify the spillover effects of learning. Second, we look at what happens if there is a change in firm’s risk attitude. The results provides explanation to the rise of imitative blockbusters in the last decades.

For all the counter-factuals, we introduce the changes at the steady state of the model. For the model to have a steady state, we remove the demand trend and set both the market size and ticket price constant over time. The rest parameters are set at their estimates. To reach the steady state, the model is simulated for a long enough “burn-in” period. We check across the paths from several independent simulations to make sure that they do converge to the same state.

7.1 Learning

What happens to a firm if it starts ignoring the demand correlation across products? Can the whole industry do as well as before? How important is the learning by the other firms vs. the learning by oneself? Our first set of counterfactuals are designed to answer these questions. We first examine the case where a single firm stops learning, which is illustrated in the top row of Figure 7. The industry is at steady state at the beginning of the plotted period. Starting from the tenth year, Firm 1 (corresponding to a major studio) treats $\lambda = 0$. The plots are averaged over multiple paths that are simulated independently.

There are several predictions. First, the firm invests in slightly fewer movies per year. Second, the firm shifts towards smaller-budget movies. This is because the absence of learning means that the firm faces a much larger uncertainty in each candidate’s $\xi$, which makes it avoid big-budget movies. In addition, we see a sizable decrease in the average latent quality. This is because without learning, the firm becomes less effective in selecting better movies. The decreases in
Each plot shows how one of the following statistics change over time: (i) the number of movies produced by Firm 1 (corresponding to a major studio) in each year, (ii) the average latent quality of these movies, (iii) the average budget size of these movies, and (iv) the average number of precursors of these movies. The solid line is averaged over many independently simulated paths. The dashed lines represent the ±2 standard deviations of the statistic across those paths. For each path the simulation starts long before time 0 to reach steady state. For the top row, firm 1 stops learning (treats λ as zero) after Year 10. For the middle row, all the firms except firm 1 stop learning. For the bottom row, all firms stop learning.

Figure 7: What Happens If Firms Stop Learning
budget size and latent quality together suggest a decline in the profitability for the studio as well as the consumer welfare.

It is interesting to compare these predictions with those where the other firms but Firm 1 stop learning, which are displayed in the middle row of Figure 7. The subjects of the plots are still the movies of Firm 1. Nevertheless, we still see decreases in the production rate, budget size and latent quality. This is exactly because of social learning. As the other firms stop learning, the movies they make become smaller and of lower quality. This means that Firm 1 will receive smaller-budget and lower-quality candidates. Although the decreases are much more gradual, the eventual sizes of the decreases are comparable to those in the first counterfactual. This suggests that a single firm benefits substantially from the learning by the other firms. This holds true in our model even though there are significant barriers against social learning, in the sense that a studio has a favorable position in imitating its own movies.  

Finally, we want to include the case where all the firms stop learning. This is displayed in the bottom row of Figure 7. As expected, we see much larger decreases in budget size and latent quality. Notice that after the initial drop, there is a gradual increase in the number of movies produced. This is caused by the larger decrease in the budget size, which, combined with the mis-perceived novelty in the candidates, makes the firms see high profitability (Figure 5) and accept more candidates.

7.2 Risk Aversion

Given the important role of risk aversion in our model, we now turn to examine what happens if there is a change in the level of risk aversion. Changes in the level of risk aversion could be caused by factors such as the risk attitude of the studio managers (Lambert (1986)), the diversification of the parent company, or more broadly the condition of the financial markets. Figure 8 displays the scenario where the coefficient of risk aversion for all firms decreases by 20% and stays at that level thereafter. Again, the plots are averaged across multiple paths that are simulated independently.

As expected, there are increases in both the number and budget size of the movies produced each year: as the firms become less risk averse, they accept more candidates. In particular, they become more accepting towards bigger-budget candidates, as they involve higher level of risks than their low-budget counterparts. The releases of these movies pave the way for further increase in the average budget, as they create the role models that the production of big-budget movies acutely requires.

There is a noticeable increase in the average number of precursors, which may come as a surprise: despite of firms being less risk-averse, movies become more imitative on average. To understand

23 Though in this paper we are looking at firm learning about consumer demand, the findings echo the literature of learning-by-doing spillovers. See, for example, Argote and Epple (1990), Irwin and Kelnow (1994), Benkard (2000) and Thornton and Thompson (2001).
The plot definitions follow Figure 7. The statistics plotted in the bottom row are: (i) the number of small & novel movies made in each year, (ii) the number of small & imitative movies made in each year, (iii) the number of big & novel movies made in each year, and (iv) the number of big & imitative movies made in each year. A movie is classified as small if its budget < $15 million, and as big if its budget is > $65 million. A movie is classified as novel if the number of precursors ≤ 2, and as imitative if the number of precursors ≥ 9. The coefficient of risk aversion decreases by 20% in Year 10 for all the firms and remains at that level thereafter.

Figure 8: What Happens If Firms Become Less Risk Averse

Here plotted are the data. The plotted statistics are the same as those in the bottom row in Figure 7.

Figure 9: Rise of Big-Budget Imitative Movies
this rise of imitativeness, we want to draw attention to several factors driving the degree of imitation. First, for any fixed size of budget, lower risk aversion implies that firms are more accepting towards original movies. This factor tends to increase the level of novelty and seems the most intuitive. However, there are two other less obvious factors working in the opposite direction. One is that with more movies being produced each year, the overall level of novelty of the arrivals decreases. Put differently, a large pool of existing movies makes it difficult to come up with something original.

The other factor comes from the fact that big-budget movies heavily rely on imitation (Figure 5). Lower risk aversion allows firms to produce the bigger-budget movies, but for these movies imitation is still required to keep the risks acceptable. It is instructive to break the movies into four categories by budget size and imitativeness, and see how the size of each category changes. This is plotted in the bottom row of Figure 8. Most noticeably, there is a large increase in the number of big & imitative movies. So it is really the infusion of a population of big-budget and imitative movies that increases the overall imitativeness.

We can compare the simulation with data. Figure 9 plots the size of each category since 1990 in the data. The data look more volatile. This is partly because in the counter-factuals the plots are averaged over multiple paths. In addition, there probably have been more events than just a single change in the level of risk aversion. Nevertheless, you can see that the most salient feature is a big increase in the number of big & imitative movies, while the changes in the other categories are less obvious. This is very consistent with our model simulation.

There have been discussions and debates in media on the rise of imitative and big-budget movies. Some commentator observes that the movie business model becomes increasingly reliant on “blockbusters – especially sequels and franchises.”24 Spielberg and Lucas, among others, expressed concerns over the declining originality in motion pictures, and point the finger at risk-focused studios.25 However, our counter-factual suggests that one probably should view the trend as an indicator that the studios have become less risk averse.

As a matter of fact, in 1989 and early 1990s, a series of conglomerate purchases and mergers that happened in the motion picture industry brought several studios new financial capabilities. It has also become more popular to co-finance movies since the 1990s.26 Both can be seen as factors that lower the level of risk aversion. Lower risk aversion makes the blockbusters possible, but does not necessarily imply a setback in the production of novel movies. This is seen in the data (Figure 9). In addition, we should remember that there are more and more movies made in each year, and this makes the creation of original work more and more difficult.

26: Co-financing is not explicitly modeled in this paper. Interested readers may look at Goettler and Leslie (2005).
8 Concluding Remarks

This paper studies product entry in the presence of firm learning from the demand for previous products. We make novel use of the data from IMDb and Amazon to construct a network amongst the products. The network allows us to examine the correlation across the market outcomes for different products. Interpreting it as a similarity network, we are able to measure the imitativeness of different products. We demonstrate the use of evolving networks in modeling industry dynamics. The model, together with its estimates, allows us to quantify the effects of learning, generate insights into the balance of innovation vs. imitation, and examine the role of risk aversion. Given our focus on the U.S. motion picture industry, it is natural to ask how our study can be extended and generalized.

In many aspects, TV shows and book publishing are similar to the motion picture industry. Series borrow each other’s scenes and novels reinvent each other’s characters. Moreover, there is probably substantial learning across these industries. It would be interesting to see how the product successes in contiguous industries lead to adaptations. In the smart-phone app market, millions of applications are being developed and distributed. Some examples of imitation are quite noticeable (e.g., Uber vs. Lyft). The strong substitutability between similar apps means that competition should be an important factor considered by app developers. It would be interesting to see to what extent apps differentiate vs. imitate each other. Fashion design is well known for being forbearing about imitation. While also present in other industries, the influence of products on consumer tastes is perhaps particularly strong in fashion design (Pesendorfer (1995), Raustiala and Sprigman (2006)). Instead of only passively learning about the consumer demand, firms can actively shape it.

The wide spectrum of budget size is also seen in scientific research. People debate about the emergence of big science, the research that is expensive and usually involves large teams. In line with our insight on the balance between innovation and imitation, while some projects have reached the stage for large scale, new projects “whose exact nature is unpredictable” are perhaps better carried out as small science (Alberts (2012)). In science, there is an alternative yet natural way to construct networks. Citation networks have been widely studied, but not much emphasis has been put on learning and formation dynamics, where important insights are conceivable. For example, one peculiar feature of science is that we use the amount of imitation, usually the citation count, to measure how successful a paper is. This may help create a cluster of research that thrives on its own - a citation “bubble” (Schmidhuber (2011)).
9 Appendix

9.1 A Toy Model

Consider a simple model of product entry with learning. There is one single firm and potential products arrive at a Poisson rate $\eta$. There is no production time, so if the arrival is accepted it gets released and generates revenue immediately. Now suppose that at time $t$ there arrives a potential product. Let us temporarily label this product by $j$. It is randomly assigned to be similar to one of the $n$ last released products: $j-1$, $j-2$, ..., $j-n$. The products older than $j-n$ become obsolete and are not imitated anymore. Let $y(j)$ denote the product that $j$ is similar to. The log return of $j$, $\xi_j$, is drawn from a normal distribution $\mathcal{N}(\lambda \xi_{y(j)}, \sigma^2)$, where $\lambda \in (0, 1)$. At time $t$, the firm does not observe $\xi_j$ but knows $y(j)$ and $\xi_{y(j)}$, so that its expectation on $j$’s log return is $\bar{\xi}_j \equiv \lambda \xi_{y(j)}$. Let $z_j$ be a product-specific cost shock known to the firm but not to us. We assume that $z_j \sim \mathcal{N}(0, \rho^2)$. The firm accepts $j$ iff $\xi_j - z_j > 0$, and discards $j$ otherwise.

The model has five parameters: $\eta$, $n$, $\lambda$, $\sigma$ and $\rho$. The question is whether we can identify all of them. The answer is yes. Technically, the identification works as follows. Let $A$ be the set of accepted products within a period of length $T$. First, parameter $n$ can be simply identified with $\max_{j \in A} |j - y(j)|$. Next, noticing that $\nu_j \equiv \xi_j - \lambda \xi_{y(j)}$ is zero-mean normal with variance $\sigma^2$ and is i.i.d. across the accepted $j$, we can identify $\lambda$ and $\sigma$ by simply regressing $\xi_j$ on $\xi_{y(j)}$ for $j \in A$. Next, because a smaller $\rho$ makes the firm more selective in accepting products, the average expected log return of the accepted product, $\frac{1}{|A|} \sum_{j \in A} \bar{\xi}_j$, can be used to identify $\rho$. In the extreme case $\rho = +\infty$, there is no selection and the average should converges to 0. Finally, given $n$, $\lambda$, $\sigma$ and $\rho$, the production rate $|A|/T$ is strictly increasing in the arrival rate so it can be used to identify $\eta$.

When there are additional revenues or costs that are proportional to the ones used in calculating $\xi$, we can model them by adding an intercept parameter to the firm’s decision. A product is accepted iff $\bar{\xi}_j - z_j - c > 0$. We want to ask if $c$ can be identified. From econometrician’s perspective, the acceptance probability is

$$\Pr(j \text{ is accepted}) = \Psi \left( \frac{\bar{\xi}_j - c}{\rho} \right) \approx \Psi(-c/\rho) + \frac{\psi(-c/\rho)}{\rho} \cdot \bar{\xi}_j,$$

where $\Psi (\psi)$ is the cdf (pdf) of the standard normal distribution. The second line is a linear approximation of the probit model around $\bar{\xi}_j = 0$. Now consider another set of parameters $(\eta', n', \lambda', \sigma', \rho', c')$ where $n' = n$, $\lambda' = \lambda$, $\sigma' = \sigma$, $c' = 0$ and

$$\rho' = \frac{\Psi(-c/\rho)}{\Psi(0)} \cdot \frac{\psi(0)}{\psi(-c/\rho)} \times \rho.$$
The acceptance probability becomes:

\[ \Pr(j \text{ is accepted})' \simeq \Psi(0) + \frac{\psi(0)}{\rho'} \cdot \xi_j \]

\[ \simeq \frac{\Psi(0)}{\Psi(-c/\rho)} \times \Pr(j \text{ is accepted}). \]

In other words, the acceptance probability is \( \varphi(0)/\varphi(-c/\rho) \) times larger than before for every arrival product. If we choose the arrival rate

\[ \eta' = \frac{\Psi(-c/\rho)}{\Psi(0)} \times \eta, \]

then the two sets of parameters are observationally equivalent. So had we specified a linear probability model instead of probit, parameter \( c \) would not be identified.

### 9.2 Details on the Attachment

Fix a point of time \( t \), the set of existing nodes and their network \( Y \). The arriving node is \( j \). Let \( p_{k,j} \) be the 1st-stage attachment probability between \( j \) and an existing node \( k \). The probability that there will be no link between \( j \) and \( k \) after the two-stage attachment process is

\[ 1 - \Pr(y_{k,j} = 1|F_t) = (1 - p_{k,j}) \cdot \prod_{\ell \sim k} (1 - p_{\ell,j} \omega), \]

where \( \ell \sim k \) indicates that \( \ell \) and \( k \) are linked in \( Y \).

In principle, given the value of \( \Pr(y_{k,j} = 1|F_t) \) for all \( k \) (as specified by (2)), one could solve for the \( p_{k,j} \)'s. However, it poses a big computational burden to solve a nonlinear system with thousands of equations every time an arrival needs to be simulated. One heuristic approach is to seek approximate solutions by postulating that \( p_{k,j} \simeq p_{\ell,j} \) for \( k \sim \ell \). Given that the network features many layers of homophily (firm, release time, budget, latent quality), it does not seem an unreasonable assumption. In this case,

\[ 1 - \Pr(y_{k,j} = 1|F_t) \simeq (1 - p_{k,j})(1 - p_{k,j} \omega)^{d_k(Y)}, \]

where \( d_k(Y) \) is the degree of \( k \) in \( Y \), i.e., the number of links connecting \( k \). Taking the log of both sides, we have

\[ \log [1 - \Pr(y_{k,j} = 1|F_t)] \simeq \log(1 - p_{k,j}) + d_k(Y) \log(1 - p_{k,j} \omega). \]

Given that the attachment probabilities are generally small (less than 1%), we may use the
first-order Taylor approximation of log to obtain

\[- \Pr(y_{j,k} = 1 | F_t) \simeq -p_{k,j} - d_k(Y)p_{k,j}\omega,\]

which implies

\[p_{k,j} \simeq \frac{\Pr(y_{j,k} = 1 | F_t)}{1 + \omega d_k(Y)}. \tag{6}\]

The denominator evens out the 2nd stage’s added attachment probability to nodes with higher degrees. We use (6) to readily compute the 1st-stage probabilities.

We can make a comparison with the alternative specification where one uses the right hand side of (2) directly as the 1st-stage probabilities. Such a specification implies that nodes with higher degrees are more likely to be attached to, similar to the concept of preferential attachment (Barabási and Albert (1999)). This leads to two undesirable features in our context. First, the probability of an original arrival (without precursors) is invariant to the density of the existing network. However, a sparse network implies diverse products, which should leave less room for innovation. Second, the model can become non-ergodic as a single product keeps being attached to over time. By connecting to the new entries, the product reinforces its probability of being attached to despite time discounting.

### 9.3 Details on Posterior Computation

For the exposition of this subsection we will fix a time \(t\). We use \(R\) to denote the set of released movies: \(\{k : r_k < t\}\), and \(Q\) the set of the yet to be released movies: \(\{k : a_k \leq t, r_k \geq t\}\). We use \(k \sim j\) to indicate that \(k\) and \(j\) are linked in the network. In our model the entire path up until \(t\) consists of \(F_t\) and \(\xi_Q\). It is not difficult to see that the probability of the entire path up until \(t\) can be written as

\[\Pr(F_t, \xi_Q) = \Psi(F_t) \cdot \prod_{k \in Q \cup R} \Pr(\xi_k | y_k, S_{a_k}).\]

The product term includes the arrival probabilities of the latent qualities. \(\Psi(F_t)\) is the part that includes the probabilities of the Poisson arrivals, attachments, budget sizes and production decisions. Most importantly, all these do not involve the latent qualities of the yet to be released movies, hence \(\Psi\) is a function of \(F_t\) only.

Given the specification in (3), the product term in the last equation is a joint density of the latent qualities that depends on the similarity network, start dates and release dates of the movies. Because these are included in \(F_t\), we can write

\[g(\xi_{Q \cup R}; F_t) \equiv \prod_{k : a_k \leq t} \Pr(\xi_k | y_k, S_{a_k}).\]
Then by the definition of conditional density, we have

\[
\Pr(\xi_Q | \mathcal{F}_t) = \Pr(\mathcal{F}_t, \xi_Q) \cdot \left[ \int \Pr(\mathcal{F}_t, \xi_Q) \, d\xi_Q \right]^{-1}
\]

\[
= \Psi(\mathcal{F}_t) g(\xi_Q, \mathcal{F}_t) \cdot \left[ \Psi(\mathcal{F}_t) \int g(\xi_Q, \mathcal{F}_t) \, d\xi_Q \right]^{-1}
\]

\[
= g(\xi_Q | \xi_R; \mathcal{F}_t).
\]

Given the specification in (3), one representation of the unconditional density \( g \) is

\[
\xi_k = \sum_{\ell, a_\ell \leq t} W_{k\ell} \xi_\ell + v_k.
\]

where \( v_k \sim \mathcal{N}(0, V_{kk}) \). \( W \) is a square matrix of the size \( \#\{k : r_k < t\} \), and \( V \) is a diagonal matrix of the same size. Their nonzero entries are:

\[
W_{k\ell} = \frac{\lambda |r_k - r_\ell|}{1 + \lambda \sum_{k \sim \ell, a_k < a_\ell} \varphi |r_k - r_\ell|}, \text{ if } \ell \sim k \text{ and } a_\ell < a_k,
\]

\[
V_{kk} = \frac{\sigma^2}{1 + \lambda \sum_{k \sim \ell, a_k < a_\ell} \varphi |r_k - r_\ell|}.
\]

In matrix form we can write in the matrix form

\[
\xi_Q = W_{QR} \xi_R + W_{QQ} \xi_Q + v_Q,
\]

or

\[
\xi_Q = (I - W_{QQ})^{-1}(W_{QR} \xi_R + v_Q).
\]

This tells us the distribution of \( g(\xi_Q | \xi_R; \mathcal{F}_t) \). So

\[
\Pr(\xi_Q | \mathcal{F}_t) = \mathcal{N}((I - W_{QQ})^{-1}W_{QR} \xi_R, (I - W_{QQ})^{-1}V_{QQ}(I - W_{QQ}')^{-1})
\]

To calculate the posterior on any subset \( O \subseteq Q \) we can simply embark the calculation of the posterior on the entire \( Q \). However, many times this is unnecessary and adds significant computational time in estimation or simulation because a large number of posteriors need to be calculated. In fact, \( g \) belongs to the class of Gaussian Markov Random Field (Rue and Held (2005)), where two sets of nodes are conditionally independent given the values of a third set of nodes if the the third set separates the first two sets, i.e., every path connecting the two sets uses nodes in the third set.

By this result, one can show that the above equation still holds if we replace \( Q \) with the collection of the nodes in \( Q \) that are not separated from \( O \) by \( R \), and replace \( R \) with the collection of the nodes in \( R \) that are directly linked to some node in the replacement of \( Q \). In the special case where \( O \) is the single arrival movie \( j \) and it is only linked to already released movies, the equation

35
Table 7: Monte Carlo Experiments for Supply Estimation

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Percent Bias</th>
<th>Percent Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attachment</td>
<td>1.0</td>
<td>1.2</td>
</tr>
<tr>
<td>2nd-stage Probability ($\omega$)</td>
<td>0.7</td>
<td>0.4</td>
</tr>
<tr>
<td>Intercept ($\gamma_0$)</td>
<td>0.7</td>
<td>3.5</td>
</tr>
<tr>
<td>Own Movies ($\gamma_1$)</td>
<td>0.7</td>
<td>2.0</td>
</tr>
<tr>
<td>Time Difference ($\gamma_2$)</td>
<td>0.5</td>
<td>2.0</td>
</tr>
<tr>
<td>Obs. Characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Budget Mean (No Precursors) ($\mu$)</td>
<td>4.2</td>
<td>12.7</td>
</tr>
<tr>
<td>Budget Variance-to-Mean ($\chi$)</td>
<td>-2.1</td>
<td>3.2</td>
</tr>
<tr>
<td>Coeff. of Risk Aversion ($\alpha$)</td>
<td>3.6</td>
<td>16.3</td>
</tr>
<tr>
<td>Shock Size ($\rho$)</td>
<td>0.7</td>
<td>4.8</td>
</tr>
<tr>
<td>Yearly Arrival Rate ($\eta$)</td>
<td>-0.3</td>
<td>5.9</td>
</tr>
</tbody>
</table>

The model is simulated from 1995 to 2012 conditional on the data from 1975 to 1994. Parameters are set equal to their point estimates. Estimation is performed in the same way as on the real data, except that it treats demand parameter values as known. The experiment is repeated for 16 times. The first column shows the average bias of the estimate for each parameter, as percentage of the absolute value of the parameter. The second column shows the standard deviation of the estimates for each parameter, as percentage of the absolute value of the parameter.

reduces to (3), the arrival distribution of $j$.

9.4 Monte Carlo

We use Monte Carlo experiment to assess the supply-side estimator. The exercise consists of simulating the model under the parameter estimates to generate a dataset with the size similar to our real sample, and then applying the supply-side estimator to the dataset to recover the parameter values. We repeat this exercise a number of times to evaluate the distribution of the estimator. The results are displayed in Table 7. All the parameters are recovered with absolute bias smaller than 5%. The last column displays the dispersion of the estimator. The standard deviations are used as the parametric bootstrapping standard errors for the supply-side estimates (see Table 6).
References


