1-1-2016

Essays on Empirical Industrial Organization and Networks

Yanhao Wei
University of Pennsylvania, yanhao@sas.upenn.edu

Essays on Empirical Industrial Organization and Networks

Abstract
This dissertation is composed of two essays in the field of empirical industrial organization. They both examine how network structures arise in and affect markets. I focus on two industries. The first one is the airline industry, and the second one is the motion picture industry.

The first essay (chapter) studies airline networks. Airlines often match higher pas-

senger density with higher flight frequency. Meanwhile, a higher frequency reduces schedule delays, creating better service quality. This suggests that, on airline networks, the value of a link to passengers increases with the density on that link. I estimate a discrete choice model for U.S. airlines with endogenous link density. The model allows me to account for changes in frequencies in counterfactual experiments. I derive implications for airline pricing, market concentration and hub-and-spoke networks.

The second essay studies product entry in the presence of firm learning from the market outcomes of past products. Focusing on the U.S. motion picture industry, I construct a network capturing the similarity amongst the movies released in the last decades. I 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. I demonstrate that learning is an important factor in entry decisions and provide insights on the innovation vs. imitation tradeoff. In particular, I find that one firm benefits substantially from the learning by the other firms. I 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.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Economics

First Advisor
Holger W. Sieg

Keywords
Airline industry, Motion picture industry, Networks, Structural estimation

Subject Categories
Economics

ESSAYS ON EMPIRICAL INDUSTRIAL ORGANIZATION AND NETWORKS

Yanhao Wei

A DISSERTATION

in

Economics

Presented to the Faculties of the University of Pennsylvania

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

____

Holger Sieg,
Professor of Economics

Graduate Group Chairperson

____

Jesus Fernandez-Villaverde,
Professor of Economics

Dissertation Committee:

Holger Sieg (Chair), Professor of Economics
Eric Bradlow, Professor of Marketing
Joseph Harrington, Professor of Business Economics and Public Policy
Katja Seim, Associate Professor of Business Economics and Public Policy
Christophe Van den Bulte, Professor of Marketing

2016

Yanhao Wei

ACKNOWLEDGMENTS

I want to thank my adviser Holger Sieg, and the other members of my committee, Eric Bradlow, Joseph Harrington, Katja Seim and Christophe Van den Bulte, for their guidance and support during the years that I worked on this dissertation. In addition, my research has benefited from conversations with many other faculty members, among which I want to express particular gratitude to Xun Tang and Pinar Yildirim. I also want to thank Editor Andrew Sweeting at the Journal of Industrial Economics, who provided detailed and helpful comments to revise an earlier version of the first chapter of this dissertation. Special thanks go to my fellow students, Xiang Fang, Qing Gong, Ju Hu, Hui Li, Yang Liu, Qiusha Peng, Devin Reilly, Hongxun Ruan, Francisco Silva and Weilong Zhang for their companionship. The financial support from the Teece Dissertation Research Fellowship awarded by the University is gratefully acknowledged.

Finally, I thank my parents for their support and encouragement throughout my education, which began when I was a baby.

ABSTRACT

Yanhao Wei

Holger Sieg

This dissertation is composed of two essays in the field of empirical industrial organization. They both examine how network structures arise in and affect markets. I focus on two industries. The first one is the airline industry, and the second one is the motion picture industry.

The first essay (chapter) studies airline networks. Airlines often match higher passenger density with higher flight frequency. Meanwhile, a higher frequency reduces schedule delays, creating better service quality. This suggests that, on airline networks, the value of a link to passengers increases with the density on that link. I estimate a discrete choice model for U.S. airlines with endogenous link density. The model allows me to account for changes in frequencies in counterfactual experiments. I derive implications for airline pricing, market concentration and hub-and-spoke networks.

The second essay studies product entry in the presence of firm learning from the market outcomes of past products. Focusing on the U.S. motion picture industry, I construct a network capturing the similarity amongst the movies released in the last decades. I 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. I demonstrate that learning is an important factor in entry decisions and provide insights on the innovation vs. imitation tradeoff. In particular, I find that one firm benefits substantially from the learning by the other firms. I 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.

Contents

List of Tables

List of Figures

Preface

Within economics, the field of industrial organization is concerned with the structure of industries and the behavior of firms and individuals in these industries. In terms of empirical work, the most active strand of empirical research into the 1980s consisted of cross-industry regression analyses, which has its roots in the "structure-conduct-performance" paradigm. Unfortunately, most studies in this literature lacked convincing econometric strategies for the identification of causal effects. This concern, together with the development of clearer theoretical foundations for analyzing imperfect markets set the stage for a dramatic shift in the 1980s towards "New Empirical Industrial Organization." The new wave of research set out to understand the institutional details of particular industries and to use this knowledge to write down and estimate models that can be used for counterfactual analysis. The findings are then generalized to other contexts, and the developed empirical methods can often be applied in multiple settings.

The study of networks has emerged in diverse disciplines, for example, physics, biology, computer science, and psychology, as a means of analyzing systems with interdependent components. The earliest known paper in this field is the famous Seven Bridges of Königsberg written by Euler in 1736, which led to the development of graph theory. In the 1930s, psychologist Jacob Moreno used a network to represent the friendship between schoolchildren, which laid the foundation for the study of social networks. Recently, social networks have become the focus of widespread attention in economics and business research. In contrast, the limited attention given to product networks is perhaps surprising, given their economic importance. Understanding the networks of "things" is particularly relevant for industrial organization, whose basic objects of analysis go beyond consumers to encompass products and firms.

In this dissertation, I attempt to help fill this gap. My dissertation is composed of two essays (chapters). Each studies one specific industry, and pays particular attention to how network structures arise in and affect the markets. The first essay studies the airline industry, where the network structure is made of cities and the airways that connect them. The second essay studies the motion picture industry, where the network represents the similarity relations between movies. In the first essay, the products that form the basis of the analysis are the "links" of the network. In contrast, in the second essay, the products are the "nodes" of the network. In each case, the products are interdependent and that interdependence is captured by the network structure. For the airline network, I pay more attention to the *implications* of the observed network structure, and use those implications to reason why the observed network has emerged in the first place. For the movie network, I focus on the mechanism behind the emergence of the network, and explicitly model its *formation* dynamics.

Chapter 1

Demand on Airline Networks

Abstract

Airlines often match higher passenger density with higher flight frequency. Meanwhile, a higher frequency reduces schedule delays, creating better service quality. This suggests that, on airline networks, the value of a link to passengers increases with the density on that link. I estimate a discrete choice model for the U.S. airlines with endogenous link density. The model allows me to account for changes in frequencies in counterfactual experiments. I derive implications for airline pricing, market concentration and hub-and-spoke networks.

1.1 Introduction

Airline carriers in the U.S. operate large-scale networks. In the network of a carrier, cities are linked by direct flights. The flux of passengers, or the "link density," is mostly satisfied with an appropriate flight frequency.[1] On the other hand, a higher frequency

[1] See, for example, Givoni and Rietveld (2009). They compare the two means to meet demand: increasing aircraft size or frequency. They find that carriers give priority to frequency.

reduces passengers' schedule delays and creates better service quality.[2] This suggests that the value of a link to passengers, and more generally the value of the network, increases with the number of passengers on that link or network. The implication coincides with the more general notion of "network effects."

There have been increasing discrete choice applications in the airline industry to examine its demand and cost structure and conduct policy analysis (e.g., Berry et al., 2006; Peters 2006; Armantier and Richard, 2008; Berry and Jia, 2010). However, researchers seem to have missed the opportunity to study in greater detail the relation between demand and link density. For example, because any decrease in demand can reinforce itself by lowering densities, the eventual impact of a higher price on demand should be larger than that in a standard model without network effects. This has implications on airline pricing and should affect the cost estimates in the BLP framework (Berry et al., 1995). More importantly, in a counterfactual analysis, one should probably account for the changes in frequencies and consequent impact on demand, instead of treating frequency as a fixed product characteristic.

In Section 1.2 of this chapter, I develop a discrete choice model that incorporates the relation between density and demand. In the model, potential passengers in a city-pair market choose which route to fly. A route may consist of one or more segments. In a carrier's network, a link can be used as a segment for multiple routes in different markets, and the link density equals the aggregate demand for these routes. On the other hand, the demand for a route is affected by the densities of its segment(s). For example, averse to empty planes, US Airways probably sets the flight frequency between Philadelphia and Los Angeles based on how many people actually fly on the link. Meanwhile, a passenger is more willing to fly on the link if the frequency is higher, making at least a part of her itinerary more smooth. Given this two-way relation, I model demand naturally as a fixed point.

[2] In the transportation literature, the demand impact of flight frequency has been well known as the "S-curve" effect. For more recent studies, see Wei and Hansen (2005) and Hansen and Liu (2015).

In Section 1.3 and 1.4, I estimate the model with data from the Airline Origin and Destination Survey (DB1B). The endogenous variables include density as well as price. Following previous applications, I regard the network structure as exogenous and construct instruments from it.[3] On the demand side, the estimation is conceptually similar to the estimation of peer effects through social networks (Bramoullé et al., 2009). It makes use of the hub-and-spoke structure of airline networks to resolve the potential "reflection" problem (Manski, 1993). On the cost side, marginal costs are inferred from the assumption of Bertrand-Nash equilibrium under the fixed-point formulation of demand.

The estimates indicate significant effects of link density on demand. This can be seen from the estimated price elasticities. A one percent increase in the overall price level is estimated to cause 1.19 percent initial decrease in the aggregate demand, close to what was found in previous studies. However, subsequent reductions in density will have further negative impact on demand. Taking it into account, the eventual decrease in the aggregate demand amounts to 1.89 percent. This higher price elasticity leads to higher marginal cost estimates. I compare the estimates with accounting numbers and find them reflective of the low profitability of the industry.

The model provides an opportunity to examine counterfactual experiments while accounting for changes in the flight frequencies. Meanwhile, it introduces some computational challenges because the densities are endogenous and all the city-pair markets are tied together by the network structure. In Section 1.5, I first address these challenges, then examine two counterfactual questions. First, what would happen had a low-cost carrier like Southwest exited the market? The result suggests a particularly strong association with market concentration and profitability, thanks to network effects. Second, what would happen if the network of a large carrier were separated in to two? The result suggests that a hub network works substantially better at a larger scale due to the higher densities that it creates. This could help us better understand

[3] I acknowledge that the instruments could be invalid in a "larger" model where carriers make network choices based on unobservables. See Section 1.3.2 for more discussions.

how hubbing emerged as a dominant feature of airline networks. At the end, I also discuss the potential use of the model for merger simulations.

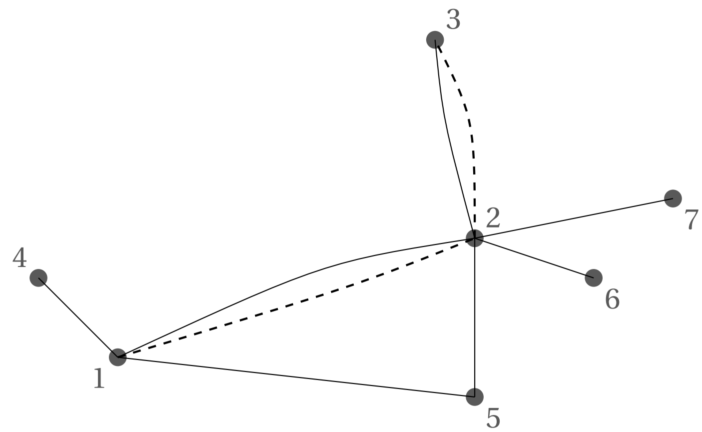
While the analysis is focused on airlines, the empirical framework in this chapter could be applicable to other industries with similar properties. For example, in retail chains, customers find it more worthwhile to visit a nearby store if it offers more items. On the other hand, the variety in the store likely depends on the size of the customer base from the nearby zip code areas. Such structure should have important implications on how retailers choose prices and store locations.

1.2 Model

The conceptual framework has carriers simultaneously set prices and flight frequencies, given long-run decisions such as market entry and hub locations. One could, perhaps, argue for a model with two stages, where carriers first set frequencies and then compete on prices. However, it should be noted that this is not required by the institutional setup: "airlines have great flexibility, even in the short term, with respect to prices, capacities, and schedules" (Norman and Strandenes, 1994). It is also common in the transportation literature to model prices and frequencies as simultaneous decisions.[4]

Underlying my model is the assumption that carriers always adjust flight frequencies to match densities. The purpose is to allow for a tractable model within the standard empirical framework for analyzing differentiated goods markets. It could give rise to richer carrier behaviors if frequency is modeled as an explicit choice variable, particularly in the two-stage game. However, the model would become less tractable

[4]For example, Dobson and Lederer (1993) look for a Nash equilibrium in which each carriers optimize flight schedule and route prices against the choices of competitors. Alder (2001) studies a game in which carriers choose networks in the 1st stage and compete on price, frequency and aircraft size in the 2nd stage. Abdelghany and Abdelghany (2010) distinguish between the strategic, planning and operation layers of airline decisions. While the strategic layer includes long-run decisions such as company expansion and hub locations, the planning layer includes flight schedule development and fare levels in each market.



Notes: The nodes represent the cities. The solid lines represent the links of carrier 1. The dashed lines represent the links of carrier 2.

Figure 1.1: An Example of Network.

because of the capacity constraints imposed by frequency choices. I offer more details on this in the appendix.

1.2.1 Links and Routes

In an airline network, a link represents direct flights by a carrier to connect two cities. Accordingly, a link is identified by a carrier and a pair of cities. I denote a link generically by g. Links are considered un-directed as it is rare for a carrier to serve only one direction between two cities (for example, flights from Philadelphia to Los Angeles but not the reverse). Figure 1.1 provides a simple example of two carriers serving seven cities, where the links of one carrier are drawn in solid lines, and the links of the other carrier are drawn in dashed lines.

Not all city pairs are connected with direct flights. In many cases passengers have to fly through multiple links to arrive at the destination. In the example of Figure 1.1, for a passenger to fly between city 2 and 4, she has to take the first carrier's one-stop route that makes a connection at city 1. More generally, a route j is a set of links $\{g, g', ...\}$ that carry passengers between two cities. These links are often called the segments of the route. A nonstop route consists of a single link, while a one-stop route is made of two links that share a common end. Routes are un-directed, as with links.

In general, the segments of a route can be operated by different carriers if there is a code-share agreement. However, in the data, such routes carried only a bit more than 2 percent of the passengers in the U.S. For simplicity, I do not consider these routes in this essay. So routes are carrier-specific. It is also rare to see passengers make detours. Accordingly, I do not consider routes that make more than one stop or whose length is more than 2.5 times the point-to-point distance. In Figure 1.1, such routes include the one flying between city 3 and 4 and the one between city 2 and 5 that makes a connection at city 1. Conceptually, it is not difficult to extend the model here to include code-shared or multi-stop routes.

Given the demand for each route, the *flow* on a link is simply the sum of the passengers across the routes that use the link as a segment (because in my analysis the time frame is fixed, flow and density are used interchangeably). Formally, let n denote the total number of routes of all carriers and D_j denote the demand for route j. Given the network structure, the flow on g is

$$F_g(D) = \sum_{j=1}^{n} D_j \cdot \mathbf{1}\{g \in j\}, \tag{1.2.1}$$

where $g \in j$ denotes that link g is a segment of j. Remember that both route and links are carrier-specific, so $\mathbf{1}\{g \in j\}$ excludes any route that does not belong to the carrier operating g. In the example of Figure 1.1, the flow on the link of the second carrier between city 1 and 2 sums the demand across two routes: i) the carrier's nonstop route between city 1 and 2, and ii) the carrier's one-stop route between city 1 and 3.

1.2.2 Demand

Fix a city-pair market m. The products in the market are the routes that serve the passengers between the two cities. They are differentiated by travel length, price, frequency, and so forth.[5] In the fashion of discrete choice models (McFadden 1981), the utility of taking route j for individual i is specified as

$$u_{ij} = \beta' x_j - \alpha p_j + f_j + \xi_j + \varepsilon_{ij}. \tag{1.2.2}$$

The first term x_j is a vector of observed characteristics of the route, such as the travel distance and whether it is non-stop route; p_j is the price; ξ_j is a route-specific fixed effect that captures econometrician-unobserved characteristics of the route; f_j measures the segment flow(s) on route j, and captures the consumer utility derived from flight frequency. For a nonstop route consisting of a single link g, I specify that

$$f_j = \theta_1 \log\left[F_g(D)\right], \qquad \text{where } j = \{g\}, \tag{1.2.3}$$

where F_g, as defined in (1.2.1), is the flow on link g. The logarithm specification is motivated by the diminishing rate of reduction in schedule delays as flight frequency increases: the average interval between two flights changes very little when the frequency is increased from 10 to 11 flights a day. Alternatively, one can think of a route as the nest of its flights as in nested logit models. In this way, θ_1 is equivalent to the "log-sum" parameter.

As to a one-stop route, the densities on both segments matter. One possible specification is a Cobb-Douglas function inside the log term:

$$f_j = \theta_2 \log\sqrt{F_g(D) \cdot F_{g'}(D)}, \qquad \text{where } j = \{g, g'\}. \tag{1.2.4}$$

[5] In some previous discrete-choice applications, markets have been defined as *ordered* city pairs and products are the directed routes or round-trip routes. While conceptually it is not difficult to apply such definitions in my model, the increased number of products will render the computation infeasible.

In this way, the smaller density matters more than the larger density. This is intuitive because the lower-density segment is more likely the bottleneck of the route. Also notice that in general $\theta_2 \neq \theta_1$. One probably expects that $\theta_2 > \theta_1$ because for one-stop routes, flight frequency also affects the waiting time at the connection.

It is possible to use a more flexible functional form to calibrate the relative importance of the larger vs. smaller density. For example, inside the log term, one can use the CES production function: $\left(\frac{1}{2}F_g^\sigma + \frac{1}{2}F_{g'}^\sigma\right)^{1/\sigma}$. As σ moves from $-\infty$ to $+\infty$, it morphs from $\min\{F_g, F_{g'}\}$ to $\max\{F_g, F_{g'}\}$. At $\sigma = 1$, it reduces to $\frac{1}{2}F_g + \frac{1}{2}F_{g'}$. At $\sigma = 0$, it becomes Cobb-Douglas. As we will see later, the point estimate of σ turns out to be close to zero, so (1.2.4) is a good approximation for the range of densities in the data.

In the model I let density enter the utility directly. A slightly different approach is to estimate a relation between density and frequency. This way, density enters the utility indirectly through frequency. However, this seems unnecessary because what matters in the end is how density affects demand. Moreover, apart from flight frequency, f could capture other factors related to density. For example, to a lesser degree, carriers also increase aircraft size to meet higher density (Givoni and Rietveld, 2009); larger aircrafts are usually more comfortable.

A distributional assumption on ε_{ij} is required to complete the demand model. As in Berry, Carnall and Spiller (hereafter BCS) (2006), I let ε_{ij} follow the distribution necessary to generate the nested logit probabilities where all the routes in the market are nested against an "outside option" of not flying. Given the network structure, denote the non-idiosyncratic part of u_{ij} by $v_j(p_j, \xi_j, D) \equiv \beta' x_j - \alpha p_j + f_j(D; \theta, \sigma) + \xi_j$. The demand for route j is given by

$$\Psi_j(p, \xi, D) = M_m \cdot \frac{S^\lambda}{1 + S^\lambda} \cdot \frac{e^{v_j(p_j, \xi_j, D)/\lambda}}{S}, \tag{1.2.5}$$

where M_m is the market size, λ is the logsum parameter, and S is a summation term across all the routes in market m:

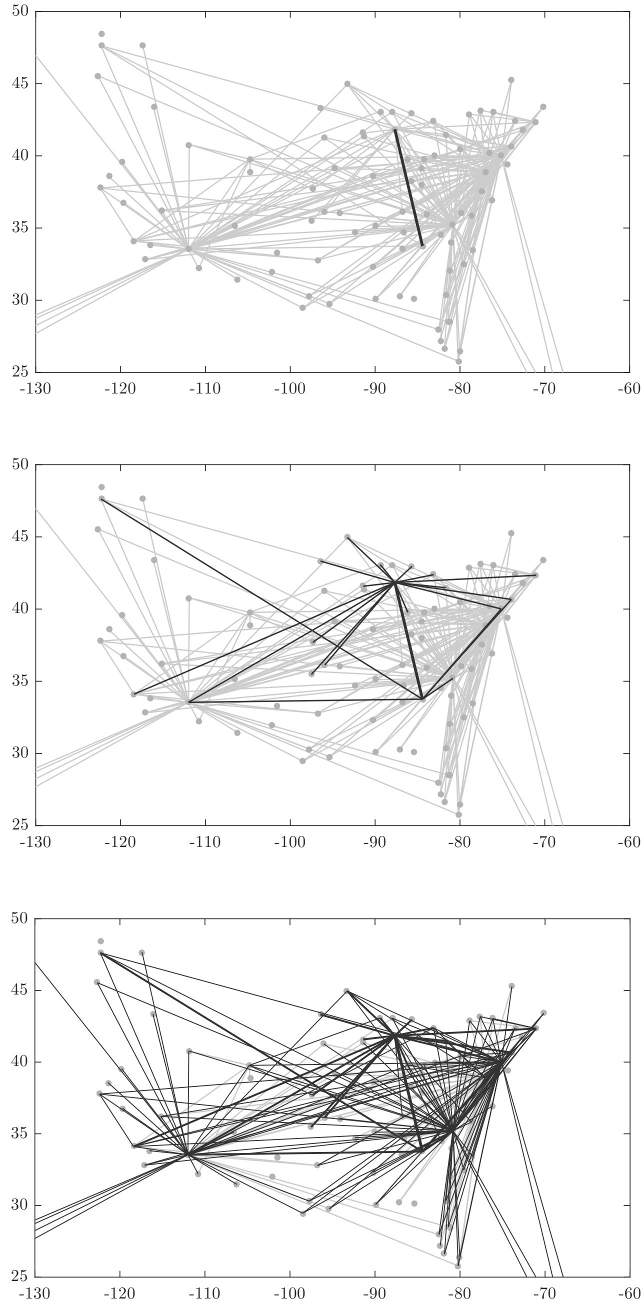
$$S \equiv \sum_{k \in m} e^{v_k(p_k, \xi_k, D)/\lambda},$$

Following BCS, market size is defined as the geometric mean of the population of the two cities in the market.

Notice that the demand vector D enters as an input of the logit model: $\Psi_j(p, \xi, D)$. This reflects network effects: the more people fly on a network, the more valuable the network becomes to consumers. The usual way of modeling network effects is using fixed points. Accordingly, I say that D is the demand predicted by the model if $D_j = \Psi_j(p, \xi, D)$ and $D_j > 0$ for $j = 1, ..., n$. In other words, it is a fixed point $\Psi(p, \xi, \cdot)$.[6] The city-pair markets are no longer independent but tied together by the network structure. To see this, it is useful to think about a price drop on a non-stop route j. First off, the market share for j will increase, leading to a higher flow on the associated link. This gives not only j further share increases but also the link-sharing one-stop routes, which are in different markets, larger shares in those markets. In general, the impact goes on and spreads over the entire network. This mechanism is illustrated in Figure 1.2 with US Airways' network. It starts with a price drop on the link between Chicago and Atlanta. Each plot shows how impact transmits from a link to another because they are the segments of a common one-stop route.[7] It is seen that almost the entire network becomes affected after three steps.

[6] I discuss the uniqueness of the fixed point in the appendix. Under certain condition on θ, an unique fixed point is guaranteed. In practice, a fixed point is found by iterating Ψ. With the estimated parameter values, the iteration always converges and I have not encountered cases where different starting points lead to different limits. In the remaining discussion of the paper, it is assumed that the fixed point is unique.

[7] Competition among routes is ignored for illustration purpose.



Notes: The plots display the sequence of impact originiated from the density change on the link between Chicago and Atlanta on US Airways' network. The impacted links are shown with a darker color. The axes are longitudes and latitudes. Competition effects are omitted.

Figure 1.2: The Spread of the Impact from a Single Link

1.2.3 Supply

Given the nature of airline operations, costs should be incurred mostly at the link level. Following BCS, I use the below specification for the marginal cost on route j:

$$mc_j = \sum_{g\in j} C(z_g, F_g; \gamma) + \omega_j, \tag{1.2.6}$$

where z_g is a vector of the observed characteristics of g, and $C(\cdot)$ can be thought as the marginal cost on link g. The last term ω_j is a route-specific fixed effect that captures unobserved determinants of marginal cost.

As to what enters C, the link length is probably the most important factor: a longer flight means more fuel consumption and more payment to the pilots. Of course, there are factors not related to flight distance, such as takeoff fuel consumption and airport landing fees. These can be captured by an intercept term. Another possible factor is the flow F_g. Economies of density (Caves et al., 1984) suggests that marginal cost is decreasing in density, or $\partial C/\partial F_g < 0$. Again following BCS, I include in C a polynomial of F_g to estimate the degree of economies of density.

To complete the model, I assume a Nash-Bertrand equilibrium in price. The equilibrium can be characterized by a set of first-order conditions. Let $D(p,\xi)$ be the fixed point of $\Psi(p,\xi,\cdot)$. For each route j that belongs to some carrier c, denoted as $j \in c$, we have

$$D_j(p,\xi) + \sum_{k\in c}(p_k - mc_k)\frac{\partial D_k(p,\xi)}{\partial p_j} = 0, \quad \text{where } j \in c. \tag{1.2.7}$$

Notice in particular that the summation is across all the routes of carrier c, not just those in the same market with j. For example, in the network of US Airways, about half of the density on the Chicago-Phoenix link comes from the direct route between the two cities, with the other half made of connecting passengers on over a hundred one-stop routes serving different markets. If US Airways invokes a price rise on that direct route, the decreased density on the Chicago-Phoenix link is going to impact those one-stop routes right away.

1.3 Estimation

The estimation builds on the general framework of Berry (1994), with two important differences. First, flow is an endogenous variable, so it needs to be instrumented aside from price. Second, demand is modeled as a fixed point, which requires adjustment in the way that we infer marginal costs from the Nash-Bertrand assumption. Below I outline the estimation procedure. The demand and supply side are estimated sequentially. In the appendix, I provide some monte carlo results, which suggest that the estimator is well behaved in finite samples.

1.3.1 Estimation Algorithm

In the data one observes the network structure, price p and demand D. Under the nested logit model, Berry (1994) provides a closed form expression for the product quality, v_j, in terms of the logsum parameter λ and the observed demand. Using that expression, one can re-write the utility specification (1.2.2) in the following way:

$$\log\left(\frac{D_j}{M_m - \sum_{k\in m} D_k}\right) = (1-\lambda)\log\left(\frac{D_j}{\sum_{k\in m} D_k}\right) + \beta' x_j - \alpha p_j + f_j(D;\theta,\sigma) + \xi_j. \tag{1.3.1}$$

The demand side can be estimated by regressing this equation. The regression is linear if $\sigma = 0$. On the right hand side, there are three terms that are endogenous in the model: $\log(D_j/\sum_{k\in m} D_k)$, p_j and f_j. They need to be instrumented. I give the details about the instruments in the next subsection.

Notice that with ξ chosen to satisfy (1.3.1), one can be assured that the observed D is a fixed point of $\Psi(p,\xi,\cdot)$. Next, to estimate the supply side, the first step is to back out the marginal costs using the Bertrand-Nash assumption. The first-order conditions (1.2.7) in matrix form is

$$D + \Delta(p - mc) = 0, \tag{1.3.2}$$

where matrix Δ is defined in the following way:

$$\Delta_{jk} = \begin{cases} \partial D_k(p,\xi)/\partial p_j & \text{if } k, j \in c \text{ for some } c \\ 0 & \text{otherwise} \end{cases}$$

To compute $\partial D_k/\partial p_j$, recall that the demand function $D(p,\xi)$ is a fixed point of $\Psi(p,\xi,\cdot)$. Provided that the usual invertibility condition is satisfied, the implicit function theorem says that $D(p,\xi)$ is locally unique in p and

$$\frac{\partial D(p,\xi)}{\partial p} = \left(I - \frac{\partial \Psi(p,\xi,D)}{\partial D}\right)^{-1} \frac{\partial \Psi(p,\xi,D)}{\partial p}. \tag{1.3.3}$$

It is not difficult to compute $\partial\Psi/\partial D$ and $\partial\Psi/\partial p$ numerically once the demand side parameters are estimated from (1.3.1). This method is fast for small networks. However, for a large and well-connected network such as the one in the data, the size of $(I - \partial\Psi/\partial D)$ is so large that the inversion becomes very difficult.

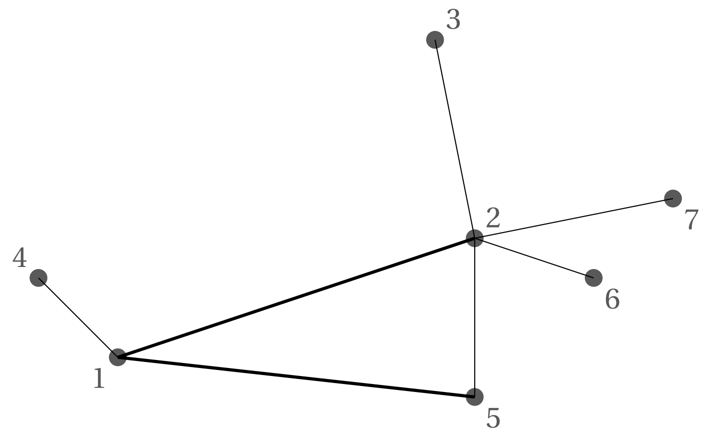
An alternative method is direct numerical differentiation. This boils down to computing the demand $D(p',\xi)$ at a perturbed price p', which can be found by iterating $\Psi(p',\xi,\cdot)$. A natural starting point of the iteration is the observed demand vector.[8] I compared the two methods on several small networks. They produced the identical results up to numerical errors. The main advantage of direct differentiation is that it does not suffer from the curse of inversion so is feasible for large networks.

Once Δ is computed, the marginal costs can be backed out as

$$mc = p + \Delta^{-1} D.$$

Notice that Δ can be arranged into a block diagonal matrix where each block corresponds to a carrier. The inversion is then carried out block by block. With the mc computed, the supply side estimation amounts to a regression of the marginal cost

[8] I tried using other starting points, but have not encountered any case where the iteration converges to a different fixed point. The convergence of the iteration is discussed in the appendix.



Notes: The nodes represent the cities; lines represent the links of a carrier. This picture is the same as Figure 1.1 except that it leaves out the second carrier.

Figure 1.3: Instruments and Identification.

equation (1.2.6). Here again I need to instrument F_g because it is an endogenous variable.

Finally, it is worthwhile to note that if one does not model D as a fixed point, treating f_j as an unchanging quantity given by the data, then equation (1.3.3) would simply be

$$\frac{\partial D(p,\xi)}{\partial p} = \frac{\partial \Psi(p,\xi,D)}{\partial p}, \tag{1.3.4}$$

which is how we normally differentiate the demand function in logit models. In other words, the term $(I - \partial\Psi/\partial D)^{-1}$ in (1.3.3) is where the notion of fixed point enters the estimation.

1.3.2 Instruments and Identification

What can be used as instruments hinges on what are exogenous in the model. Since the network structure has been treated as an exogenous primitive of the model, I derive instruments from it. These include, for example, the number of routes in a market as well as the number of routes that share a particular link. The conceptual framework is the same as in BCS and Berry and Jia (2010), where the network structure is treated as a long term choice, not subject to the unobservable shocks ξ or ω. In studies of airline entry, researchers have considered models where the shocks are unknown to carriers before entry (Aguirregabiria and Ho, 2011), as well as models where the shocks could potentially affect entry decisions (Ciliberto and Williams, 2014). In the later case, the network structure is no longer a valid source of instruments, and one is required to exploit other primitives.

More specifically, for the demand side regression (1.3.1), the number of the routes in the market is used as an instrument to estimate λ. Basically, λ is identified from the changes in the total share of the routes in a market as the number of routes varies across markets. One can think of the extreme case $\lambda \to 0$, where the total share of the routes in a market remains fixed regardless of the number of routes.

As to price, BLP suggests exploiting measures of market-level competitiveness. The more competitive a market is, the lower the prices are in a Bertrand-Nash equilibrium. Accordingly, I instrument the price on a route with the number of carriers in the market and the number of rivals' routes.

As to density, I exploit the "centrality" of links in a network. Specifically, for link g, let F_g^{IV} be the number of possible routes that include g as a segment. This instrument is relevant because a heavily utilized link is likely to bear a large flow. Figure 1.3 reproduces the example in Figure 1.1, but leaves out the second carrier for the sake of illustration. Let us compare two links: one between city 1 and 2, the other between 1 and 5. There are seven routes using the first link and only four using the second link (here for both links we count the one-stop route that serves city 2 and 5 and makes a connection at 1, though it makes quite a detour and is unlikely to be flown). Not considering competition or variations in population, it is clear that one should expect more passengers on the first link.

To better see how this instrument helps identify θ, let us continue with the example. Consider the two non-stop routes corresponding to these two links, which are similar in length. If $\theta_1 = 0$, one would expect similar demand on the two routes. However, if

$\theta_1 > 0$, there is probably a higher demand for the route between 1 and 2 because the density on the associated link is expected to be higher. The same argument applies to one-stop routes. For example, we may compare the route that serves 4 and 5 with a connection at 1, and the route that serves 1 and 6 with a connection at 2. If $\theta_2 = 0$, one would expect similar demand on these two routes. However, if $\theta_2 > 0$, there is probably a higher demand for the second route. Notice, most importantly, how we have arrived at these conclusions by only looking at the network structure, without any information about ξ or ω.

Finally, It is worthwhile to point out that the identification requires that some links are flown by multiple routes, otherwise we would have a "reflection" problem as in Manski (1993). To illustrate, consider a purely point-to-point network, in which there are no connecting routes. Because D_j and F_g coincide for each route $j = \{g\}$, the estimation essentially would regress the demand on itself. Fortunately, multi-hub networks provide abundant asymmetries between demand and density. In this regard, the demand side estimation in this chapter is conceptually similar to the work by Bramoullé et al. (2009) on peer effects. They show how social network structure can help resolve the reflection problem and be used to derive instruments for estimation. Similar techniques have also been seen in the literature of spatial econometrics (Anselin et al. 2004).

1.4 Data and Results

1.4.1 Data

The Airline Origin and Destination Survey (DB1B) is a 10% sample of airline tickets from reporting carriers in the U.S. collected by the Bureau of Transportation Statistics. I use the DB1B-coupon and the DB1B-market data for the forth quarter of 2012 and the first quarter of 2013. Some of the carriers were contracting only: they did not sell tickets to passengers but only operated flights for the major carriers. The

Table 1.1: Summary Statistics of the Networks of the Five Largest Carriers

	# Links	# Routes	# Markets served	Total flow (million)	Total passengers (million)
Southwest	573	9,070	1,721	51.8	42.6
Delta	422	10,097	3,728	43.6	31.9
United	465	11,076	3,877	31.2	24.2
American	237	5,074	2,967	26.4	20.2
US Airways	258	5,233	2,803	27.8	19.4

Notes: The networks are restricted to the top 100 cities. Compiled from DB1B data for 2012Q4 and 2013Q1.

flights of these carriers are incorporated into the carriers for which they operate.[9]

For estimation, the network includes the 100 most-visited cities and the twelve largest carriers. This includes a bit more than 87 percent of the passengers in the DB1B data. I regard a link as existing in the network iff no less than 360 passengers were observed in the DB1B data. This roughly corresponds to the number of passengers that would be carried on a weekly flight by a medium-size jet. Price p_j is the average fare observed on route j. I eliminate tickets with unusually low or high fares as they may represent coding errors. A route is included only if it there is at least one passenger observed in the data.[10] Intra-Hawaii routes are excluded. In the end, my sample includes $44{,}121$ routes and accounts for about 81 percent of the passengers in the DB1B data.

The vector of route characteristics, x_j, includes an intercept, the carrier dummies, the market distance (i.e., the point-to-point distance between the two cities in the

[9] My estimation restricts attention to the U.S. domestic market. I acknowledge that a considerable proportion of the operations of some carriers are international. However world-wide survey data on air-travel itineraries is not available.

[10] There is no passenger observed on about 12 percent of the routes that can be constructed from the network. Unfortunately, the BLP framework does not allow products with zero observed market share. Gandhi et al. (2015) suggest a simple method to rectify this problem. However, applying their method here is not straightforward, because price is unknown on a route with no observed passengers. Following the practice in the airline literature, I simply exclude these routes from estimation.

Table 1.2: Summary Statistics

	Mean	SD	Min	Max
Route level (x_j, p_j)				
Market distance (1,000 miles)	1.37	0.84	0.06	5.97
Tourism dummy	2.88	0.45	0	1
Hub dummy	0.17	0.38	0	1
One-stop dummy	0.94	0.23	0	1
Extra route length (1,000 miles)	0.35	0.36	0	3.44
Price ($100)	2.62	1.02	0.03	9.96
Link level (z_g,F_g)				
Length (1,000 miles)	0.98	0.69	0.06	4.96
Density (100,000 passengers)	0.92	1.26	0.003	25.9

Notes: There are 44,121 routes and 2,419 links. Market distance is the point-to-point distance (great-circle) between the two cities in the market. It coincides with the length of a non-stop route in the market. Extra route length is the difference between route length and market distance.

market), the square of the market distance, a dummy indicating whether either city in the market is a tourist destination, a dummy for one-stop route, the difference between the route length and the market distance, and finally, in the case that the operator is a legacy carrier, a dummy indicating whether either city in the market is a hub of the carrier. Tourist destinations include Atlantic City, Charlotte Amalie, Las Vegas, New Orleans, and places in Florida and Hawaii. The vector of link characteristics, z_g, includes the carrier dummies and the length of the link.

Table 1.1 displays some summary statistics for the networks of the five major carriers. Southwest Airlines, being the largest carrier in terms of the number of links, serves the fewest markets. This is because compared with the other carriers, it operates a more point-to-point rather than multi-hub network. Table 1.2 provides the summary statistics for some of the regression variables.

1.4.2 Parameter Estimates

1.4.2.1 Demand Side

Table 1.3 displays the estimates for the demand side parameters. In the benchmark case, the logsum parameter is estimated to be 0.56, which is not far from the estimates in BCS (0.605) or Peters (2006) (0.595). Next, notice that air-travel demand is inverted U-shaped in market distance. Short-haul flights compete with ground transportation, which become worse substitutes as distance increases, so demand initially grows with distance. As distance increases further, the gravity model kicks in and demand starts to decrease.

The price coefficient is estimated to be 0.65. This is somewhat larger than that in BCS (0.46), but their estimate was based on the data from the much earlier year of 1985. The implied aggregate price elasticity, which is the percentage change in total demand when all products' prices increase by 1 percent, is 1.19 without taking into account decreases in density and the subsequent effects on demand.[11] One can compare this to Gillen et. al. (2003) that conducted a survey of cross-sectional studies on the the elasticity of air-travel demand. The estimates ranged from 0.181 to 2.01, with a median of 1.33. So my estimate seems reasonable.

The effects of density are significant in both statistical and economic sense. At the estimated level, the aggregate price elasticity amounts to 1.89, instead of 1.19, when taking into account how decreases in density would lead to further decreases in demand.[12] Also notice that density has a larger impact on one-stop routes than non-stop routes ($\theta_2 = 0.49$ vs $\theta_1 = 0.33$). This is intuitive because on a one-stop route, flight frequency affects not only schedule delays at the arrival and departure but also the waiting time at the connection.

It is possible to test for weak instruments. In the demand regression equation there are

[11]Formally, this considers how much $\sum_{j=1}^{n} \Psi_j(\kappa \cdot p, \xi, D)$ changes when κ increases from 1 to 1.01.

[12]Formally, this considers how much $\sum_{j=1}^{n} D_j(\kappa \cdot p, \xi)$ changes when κ increases from 1 to 1.01, where $D(\kappa \cdot p, \xi)$ is the fixed point of $\Psi(\kappa \cdot p, \xi, \cdot)$.

Table 1.3: Demand-Side Parameter Estimates

	Benchmark	No IV	CES
Carrier dummies	Yes	Yes	Yes
Intercept	-2.83 (.09)	-3.33 (.03)	-2.82 (.10)
Market distance	0.28 (.02)	0.28 (.02)	0.28 (.03)
Market distance^2	-0.031 (.006)	-0.074 (.006)	-0.032 (.007)
Tourism dummy	0.32 (.01)	0.37 (.01)	0.32 (.01)
Hub dummy	0.07 (.02)	-0.17 (.02)	0.05 (.02)
One-stop dummy	-1.98 (.03)	-2.17 (.02)	-1.97 (.03)
Extra route length	-1.09 (.04)	-1.29 (.02)	-1.08 (.04)
Non-stop flow, θ_1	0.33 (.02)	0.60 (.01)	0.34 (.02)
One-stop flow, θ_2	0.49 (.01)	0.75 (.01)	0.50 (.01)
CES, σ	0	0	0.10 (.27)
Price, α	0.65 (.05)	0.19 (.01)	0.64 (.05)
Logsum, λ	0.56 (.01)	0.55 (.01)	0.57 (.01)
Cragg-Donald statistic	479.3	-	-
Function value	10.3	-	9.46

Notes: The second column differs from the benchmark case by not instrumenting for price or flow. The third column differs from the benchmark case by using the more flexible CES specification. Standard errors are constructed by bootstrapping (see Appendix).

three endogenous regressors: log within-nest share, price, and flow. When there are multiple endogenous regressors, one uses the Cragg-Donald statistic, a generalization of the first stage F-statistic (see Stock, Wright and Yogo, 2002). It is reported in Table 1.3 for the benchmark case, and indicates clearly that the instruments are not weak. The table also reports the objective function value. The J test rejects the over-identifying restrictions. However, since the model is obviously stylized and the sample is large, it should not be surprising that I could formally reject the model.

As a comparison with the benchmark case, the second column of Table 1.3 displays the estimates without instrumenting for either price or flow. Notice first that the price coefficient α is underestimated. This is consistent with a Bertrand-Nash equilibrium in which p_j and ξ_j tend to be positively correlated. Second, notice that θ_1 and θ_2 are overestimated. This is intuitive because the demand for j is increasing in ξ_j, which creates a positive correlation between ξ_j the segment densities on j.

In the third column of Table 1.3, I explore the possibility that the segment densities on a one-stop route may enter the utility in ways other than that specified in (1.2.4). The more flexible CES functional form is used. The point estimate of the substitution parameter σ is 0.10. For the range of densities in the data, this is very close to the Cobb-Douglas specification used in the benchmark case. Given the size of the bootstrapped standard error for σ, one probably should be concerned with the strength of the instruments to identify σ. Unfortunately, to my knowledge, there is not yet an established test for circumstances where a parameter enters regression non-linearly.

Finally, Table 1.5 displays, in the first column, the coefficient estimates of the carrier dummies in the benchmark case. On average, passengers find it more pleasant to fly with the legacy carriers than with Southwest, which is not surprising given Southwest's status as a low-cost carrier.

Table 1.4: Cost-Side Parameter Estimates

	Benchmark	No FP	Density	Density IV
Carrier dummies	Yes	Yes	Yes	Yes
Intercept	0.75 (.02)	0.27 (.05)	0.83 (.03)	0.69 (.12)
Link Length	0.45 (.01)	0.43 (.01)	0.45 (.01)	0.44 (.04)
Flow Density	-	-	-0.10 (.02)	-0.050 (.28)
Flow Density^2	-	-	0.014 (.005)	0.013 (.13)
Flow Density^3	-	-	-0.0004 (.0003)	0.00059 (.011)
R^2	0.300	0.264	0.305	-
Cragg-Donald stat .	-	-	-	0.17
Function value	-	-	-	0

Notes: The second column differs from the benchmark case by not imposing fixed point on demand. The third and forth column differ from the benchmark by including a polynomial of flow density. The flows are instrumented in the forth but not the third column. All columns are based on the benchmark estimates of the demand parameters. Standard errors are constructed by bootstrapping (see Appendix).

Table 1.5: Dummy Estimates for Major Carriers

	Demand Side	Cost Side
Southwest	0	0
Delta	0.20 (.02)	0.13 (.02)
United	0.32 (.03)	0.31 (.02)
American Airlines	0.14 (.02)	0.11 (.02)
US Airways	0.42 (.03)	0.24 (.02)

Notes: The coefficients for Southwest dummy of are set to zero. These estimates correspond to the benchmark cases in Table 1.3 and 1.4. Standard errors are constructed by bootstrapping (see Appendix).

1.4.2.2 Supply Side

Table 1.4 presents the cost side parameter estimates. Recall that the demand and supply are estimated sequentially. The numbers in Table 1.4 are all based on the benchmark estimates on the demand side. The standard errors are bootstrapped and take into account the errors from the demand estimation (see Appendix for details). By the specification, these estimates measure the marginal cost at the link level. According to the benchmark case, there is an estimated \$75 fixed marginal cost regardless of the length of the flight; for each additional mile, it is estimated that the marginal cost increases by 4.5 cents.

The second column displays the estimates without imposing the fixed point demand. As pointed in Section 1.3, this boils down to differentiating the demand function as in (1.3.4) instead of (1.3.3). The fixed marginal cost is estimated to be a much smaller \$27. This is because the price elasticities are higher when we account for network effects through the fixed point demand. With higher elasticities, the marginal costs that rationalize the observed prices must be higher.

One can compare the estimated costs with accounting numbers. First off, it is acknowledged that accounting practices generally are not geared toward reporting the economic notion of marginal cost. Nevertheless they still can be and have been used to complement econometric analysis (Nevo 2001, Einav and Levin 2010). Because the estimation relies on the domestic data, here I focus on Southwest, whose operations were almost exclusively domestic.[13] According to the filing of Southwest for 2013Q1, it flew 23.7 billion passenger miles and incurred \$4 billion operation costs, with \$2.1 billion spent on fuel, maintenance, aircraft rentals and landing fees, and the rest on salary, depreciation and "other operation expenses." This implies a CPM (costs per passenger mile) of 8.9~16.9 cents. The CPM based on the benchmark estimates in Table 1.4 is 12.7 cents, which seems reasonable. The CPM based on the estimates in

[13]The quarterly filings are available at http://investors.southwest.com/financials/quarterly-results/. Last access: Feb. 16, 2016.

the second column is only 5.8 cents, which seems a bit too small.[14]

To explore the possibility of economies of density, in the third column I add a polynomial of density to the regression. The first two coefficients are statistically significant. At the point estimates, the shape of the polynomial is similar to that in BCS at medium flight distance. The derivative of marginal cost with respect to density is negative on 98 percent of the links in the data. However, comparing the R^2 with that in the benchmark case, it is seen that density provides very little additional explanation to the variation in marginal cost.

Because density is endogenous, it should be instrumented and this is done in the forth column. The estimates of the polynomial coefficients are higher than those in the third column. This is intuitive because a higher ω_j often forces a higher price, leading to a negative correlation between ω_j and the segment densities on j. At the point estimates, the derivative of marginal cost with respect to density is negative on 86 percent of the links in the data. The maximum cost reduction is four and a half cents, achieved around a density of 170 thousand passengers (in two quarters).

The Cragg-Donald statistic indicates that there is not enough strength of the instruments to separately identify the coefficients of the polynomial.[15] This is also reflected in the size of the bootstrapped standard errors. One way to solve this problem is removing the higher-order terms in the polynomial. However, as argued in BCS (2006), such a specification is unlikely flexible enough to capture the complicated workings of economies of density.[16] So for the counterfactual experiments, I fall back on the benchmark model and focus exclusively on the demand side effects of density. Of course, the results must be interpreted accordingly.

The estimates of the carrier dummies for the benchmark case are displayed in the second column in Table 1.5. It is seen that the legacy carriers generally incur more

[14]Berry and Jia (2010) estimate the CPM for all the carriers to be 6 cents.

[15]Each first stage F-statistic is above 10. This shows that individual F-statistics are not sufficient when there are multiple endogenous regressors.

[16]BCS (2006) were able to estimate a polynomial of both density and distance of degree 3. But frequencies were treated exogenous and used to instrument density.

costs than Southwest. Again, this is expected given Southwest's status as a low-cost carrier.

Finally, it is worthwhile to mention that markups are *negative* on 31 percent of the routes, all of which are one-stop and make up about 6 percent of the passengers in the data. At first glance, this is a bit surprising because pricing below marginal cost on a route means that the carrier is not breaking even on that route. However, this very low price benefits the other routes in the network through increased link densities, so the carrier can still find it desirable. Put differently, the carrier prefers serving some passengers at loss rather than leaving planes empty and subsequently cutting back frequencies. This is particularly true on the connecting routes, as each contributes to the densities on more than one segment. Interestingly, the finding here matches Delta's claim that "connecting traffic is the least profitable for the airline," made in 2005 when it reduced its capacity at the Cincinnati hub.[17]

1.5 Counterfactuals

In counterfactual simulations, previous discrete choice applications in the airline industry have treated the flight frequencies as fixed quantities given by data. While it seems difficult to model them as explicit choice variables, the fixed-point model in this chapter does provide a way to account for changes in flight frequencies in counterfactuals. This offers an opportunity to re-examine or explore important questions in the airline industry. Here I pursue two.

First, there have been considerable discussions to understand the persistent financial losses of the industry (Borenstein 2011). Until 2012, the domestic airline industry had reported negative net income in 23 of 31 years since deregulation. Carriers have shown valiant efforts in creating a less hostile environment, such as maintaining "capacity discipline" and seeking consolidations.[18] This makes one wonder if there has been too

[17]"Delta to make local cuts in service, jobs." *Cincinnati Business Courier*, September 7, 2005.

[18]See Mazur (2015) for an interesting alternative explanation of the "capacity discipline."

much competition in the industry. In the counterfactual, I ask what would happen if the largest low-cost carrier, Southwest, exited the market. In particular, I compare the predictions of the fixed-point model with those of a standard model.

Second, researchers have been interested in understanding the hub-and-spoke structure that emerged after the airline deregulation. Borenstein (1989) showed that dominant airport presence provides hub premiums, which recently was further examined by Ciliberto and Williams (2010). Brueckner and Spiller (1994) focus on economies of density and provide a cost-based rationale for hubbing. The analysis in this chapter also focus on the roles of density but suggests a demand side rationale for hubbing. To see how network effects interplay with the network structure, one can examine the consequences of a merger, or reversely, a separation of a carrier's network. The results have implications for studies on the formation of airline networks (e.g., Hendricks et al. 1999, Gastner and Newman 2006).

The fixed-point model poses a few computational challenges. Because the markets are tied together through common links, tens of thousands of prices must be solved simultaneously rather than market by market.[19] Moreover, the endogeneity of the densities must be taken into account when calculating the optimal prices. Below I discuss how an equilibrium can be found within a reasonable amount of time. Readers that are not interested in computational details can go directly to Section 1.5.2.

1.5.1 Computing Price Equilibrium

Below I present an algorithm that computes a Bertrand-Nash equilibrium for a given network structure, unobservables ξ and marginal costs mc. The algorithm relies on the first-order conditions (Equation 1.3.2).

1. Choose a convergence criterion $c > 0$ and an initial value for the price vector p^*.

2. Iterate $\Psi(p^*, \xi, \cdot)$ to reach the fixed-point demand $D(p^*, \xi)$.

[19] Actually, such counterfactual exercise was thought infeasible and avoided in Berry and Jia (2011).

3. Compute $\partial D(p^*,\xi)/\partial p$ and the corresponding matrix Δ. Let $p^{**} = mc - \Delta^{-1}D(p^*,\xi)$.

4. If $\|p^{**} - p^*\|_\infty < c$, exit.

5. Updating p^* by moving it towards p^{**} and repeat step 2-5.

Step 3 is the most time-consuming for two reasons. First, it needs to differentiate an implicitly defined demand function. As discussed in Section 1.3, $\partial D/\partial p$ can be computed by either implicit function theorem or direct differentiation. Again, implicit function theorem has a speed advantage for smaller networks but seems infeasible for large networks.

The second reason is that the step needs to solve a large linear system $\Delta^{-1}D(p^*,\xi)$: the size of Δ is the total number of routes across all the carriers and markets. Notice that Δ_{jk} is generally non-zero even if j and k belong to different markets, because they may share a link, or there is a third route that shares links with both j and k, or so forth. Fortunately, $\Delta_{jk} = 0$ at least when j and k belong to different carriers. So Δ can be organized into a block-diagonal matrix with each block associated with one carrier, and the linear system can be solved block by block.

Nevertheless, a block of Δ can still be of considerable size for a major carrier such as Southwest or United. So it is important to reduce the numerical errors in Δ before solving the linear system. In this regard, I find it helpful to use the two-sided finite difference formula when computing $\partial D(p^*,\xi)/\partial p_j$. I also find it helpful to use a very small stop criterion whenever iterating Ψ for a fixed point.

As to exactly how p^* should be updated in the last step, I find that simply replacing p^* with p^{**} works well in the beginning, but dampening is often required when p^* moves close to the solution. Specifically, I replace p^* with $(1-\tau)p^* + \tau p^{**}$, where τ starts at 1 and gradually decreases as $\|p^{**} - p^*\|_\infty$ becomes close to zero. After the algorithm exits, I numerically double check whether the solution is at least locally optimal for each carrier.[20]

[20] I ran the computation with Matlab on a 64-core cluster. Computing $\partial D(p^*,\xi)/\partial p$ was run in parallel. For the counterfactuals, finding an equilibrium took roughly two hours.

Table 1.6: What If Southwest Were Not Here

	SW	Delta	United	AA	US
Benchmark model (%)					
Total passengers	-	11.3	22.8	17.3	18.3
Average price	-	0.7	0.6	0.8	0.6
No fixed point (%)					
Total passengers	-	6.1	12.4	9.7	9.2
Average price	-	1.0	0.8	1.0	0.7

Notes: All the numbers are in percentage. Average price is calculated across routes. The lower part comes from the non-fixed-point model and the associated cost parameter estimates.

1.5.2 Results

1.5.2.1 What If Southwest Were Not Here?

Table 1.6 displays the percentage changes in the demand and average price when Southwest's network is removed. The upper part reports the predictions from the benchmark model. The lower part reports the predictions from the non-fixed-point model that treats f_j as an unchanging quantity in the counterfactual experiment, with the cost side parameters set at the associated estimates (Table 1.4, second column).

In both models, the overall changes in price are small. However, it should be pointed out that if one only looks at the markets where Southwest has a significant share in the data, then the increases in price are at a much larger magnitude.

The more sizable changes are found in the demand, and it is where the predictions of the two models diverge. Qualitatively, the differences in prediction are intuitive: as Southwest exits the market, demand increases for each remaining carrier; however, only the benchmark model is able to capture the subsequent increases in flight frequencies that help these carriers acquire even more passengers. Quantitatively, for each remaining carrier, the increase in demand predicted by the benchmark model more or less doubles that predicted by the non-fixed-point model. Put in a reverse

Table 1.7: What If United's Network Were Separated into Two

	SW	Delta	United	AA	US
Benchmark model (%)					
Total passengers	2.6	2.9	-36.1	4.3	3.6
Average price	0.6	0.8	-2.7	0.9	0.5
No fixed point (%)					
Total passengers	0.6	0.8	-15.7	1.0	0.8
Average price	0.4	0.5	-1.6	0.6	0.3

Notes: All the numbers are in percentage. Average price is calculated across routes. The lower part rows come from the non-fixed-point model and the associated cost parameter estimates.

perspective, the introduction of Southwest has taken a great deal of passengers away from the legacy carriers.

As an implication, we probably should see a particularly strong association between market concentration and profitability in the airline industry, thanks to network effects. This implication is in the same direction as the one that would result from economies of density (Brueckner and Spiller 1994). It provides a new angle to understand a decade of consolidations in the industry. Most recently, US Airways CEO Doug Parker said that a merger with American was "the last major piece needed to fully rationalize the industry," meaning that the merger would consolidate the industry enough that airlines could start making profits.[21]

1.5.2.2 What If United's Network Is Separated?

In the data, United Airlines has the largest multi-hub network (see Table 1.1). Table 1.7 displays the percentage changes in the total demand and average price if we take United's network and separate it into two. More specifically, two new companies are created and each operates a randomly selected half of the links from United's original

[21] "Which Is Worse: Airline Monopolies or Airline Competition?" *Harvard Business Review*, August 15, 2013.

network. This results in the loss of 5,333 one-stop routes whose segments are not assigned to the same company, about half of the one-stop routes in United's network. The main interest here is to see the consequences of such a reduction in the scale of the hub-and-spoke structure. Conducting the same counterfactual on Delta and American Airlines turned out to produce very similar results.

The upper part of the table reports the predictions from the benchmark model, while the lower part reports the predictions from the non-fixed-point model. The total number of passengers of United after the separation adds together the passengers of both new companies; the average price averages the prices across the routes of both new companies. Again, both models predict small overall changes in price.

The non-fixed-point model predicts a 15.7 percent decrease in the total demand for United, which actually represents very closely the pre-separation demand on that 5,333 lost one-stop routes. Reasonable as it sounds, yet that prediction is more than doubled by the benchmark model, to 36.1 percent. This number is actually quite staggering, noticing that we have not removed a single link! Nevertheless, the difference in prediction is qualitatively intuitive: after the separation, the links lose the inflow of passengers from those destroyed one-stop routes; this leads to reductions in flight frequencies, which are captured only by the benchmark model.

The takeaway message is that network effects create synergy in a hub-and-spoke network, and it scales up substantially with the size of the network. Reversing the counterfactual exercise, one can conclude that a merger between the two split companies would lead to large increases in both the demand and operation profits. So this again speaks to the incentives for mergers. Of course, the caveat here is that in general the outcome of a merger depends on how many new routes can be created by the merger and how much the two pre-merger networks overlap. For example, a merger will be of little consequence if it is between two carriers that serve disjoint sets of cities, and even produce adverse consequences if it is between two completely overlapped networks.

1.5.2.3 Merger Simulation

Real mergers very often lie somewhere between the above-mentioned extreme cases, which means combining overlapping links as well as creating new routes that are not possible within either one of the two pre-merger networks. In fact, "give our customers access to more destinations than ever before" is often quoted by carriers to defend their DOJ merger cases. Accordingly, the set of products should be changed during the merger. Unfortunately, standard merger simulations with discrete choice models typically only change the ownership structure of the existent products.[22]

The model developed in this chapter uses the network structure as the sole primitive, so in principle, it can be used to simulate network mergers while accounting for changes in both the set of routes and flight frequencies. However, there remains one practical difficulty: it is not straightforward to assign values of ξ and ω to the routes on the merged network. To see this, simply consider any new route that is created by the merger, or any route with a segment at which the two pre-merger networks overlap. In my experiences, simulation results are sensitive to the specific assignment rule. I leave the merger exercise for future research.

1.6 Concluding Remarks

Two noted economists at University of Chicago, Sam Peltzman and Lester Telser, made a bet in 1979 on whether the 4-firm concentration ratio would be above 90 percent by 1985. Peltzman bet that it would not to be, and as he put it, "won the bet, but lost the war."[23] Based on the belief that no (cost side) scale economies existed at the level of major carriers, many thought that the deregulation would end airline monopoly or oligopoly. However, it is safe to say that the market has been mostly oligopolistic, probably even more so today after a decade of mergers and acquisitions.

[22] See Nevo (2000). Also see Peters (2006) on airlines.
[23] See, Borenstein (1992).

Apart from the factors such as hub premium, economies of density and entry selection (Roberts and Sweeting, 2013), this essay emphasizes that there are significant scale economies *on the demand side,* which may have driven the industry away from ideal competition.

Chapter 2

Product Similarity Network in the Motion Picture Industry

Abstract

I study product entry in the presence of firm learning from the market outcomes of past products. Focusing on the U.S. motion picture industry, I construct a network capturing the similarity amongst the movies released in the last decades. I 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. I demonstrate that learning is an important factor in entry decisions and provide insights on the innovation vs. imitation tradeoff. In particular, I find that one firm benefits substantially from the learning of the other firms. I 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.

2.1 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 chapter 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.[24] 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."[25] In fact, movie imitation frequents the media as a subject of discussion as well as debate.[26]

To better understand how firms balance innovation and imitation, I 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 essay, correlation

[24] On motion picture uncertainty, also see De Vany and Walls (1996, 1999).

[25] Squire (2005), p.4.

[26] See "Hollywood Learns Originality Does Not Pay." May 29, 2015, *Financial Times*. Also see "Are Blockbusters Destroying the Movies?" *New York Times*, Jan. 6, 2015.

is explicitly modeled. I 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. I 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, I 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 opportunities of both innovation and imitation.

I bring the model to data. To construct the network, I 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. I 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. I construct a network of nearly 4,500 movies released in the U.S. in the last decades. Through reduced-form analysis, I 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). I also find evidence that suggests firm learning and risk aversion.

The essay proceeds to estimate the model with the method of simulated moments and conduct counter-factual experiments. Several insights are derived. First, I 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. I 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, I 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. I 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, I 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.[26]

In terms of general insights, this essay adds to the empirical literature 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.[27] Goldenberg et al. (2012) study the network of hyperlinked video clips on YouTube. Oestreicher-Singer and Sundararajan (2012) examine how a product's position in a recommendation network affects its demand. Wei (2014) extends the estimation framework for differentiated markets to airline networks. This essay also adds to the literature 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 exits optimally. Toivanen and Waterson (2005), Shen and Xiao (2014) and Yang (2014) focus on how firms learn

[27]For theoretical studies on product networks, see Dellarocas et al. (2010) and Mayzlin and Yoganarasimhan (2012).

from the entry/exit *choices* of each other in the context of fast food chains. In this essay I look at a different channel of learning, namely from the market outcomes of past products. More broadly, the essay is related to the literature on learning models (Ching et al., 2013). In terms of the application, the essay 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 has not been a study that models the entry decisions where studios green light movies. This essay provides a first attempt from the perspective of learning.

The rest of the chapter is organized as follows. Section 2.2 describes the data, in particular the construction of the network. Section 2.3 presents the reduced-form analysis. Section 2.4 presents the model. Section 2.5 describes how I estimate the model. Section 2.6 presents the estimates. Section 2.7 conducts counter-factual experiments. The last section concludes and discusses future research.

2.2 Data

Data Sources I 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 I want to study firm's go/no go decisions, I 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 essay I focus on the box office at the first release. For a small number of movies whose budget sizes are missing on IMDb, I am able to collect them from Wikipedia.com. To account for inflation, I 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, I 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. I 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. The idea is to find out which 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, I 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. I 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).

I 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 chapter, I combine the IMDb and Amazon references.[28] As a robust check, I run all the analyses throughout the chapter 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.[29] Web cookies were disabled to prevent the references from being tailored for the crawling history. I collected the data at two different time points, first in March 2014 and then in October 2015. The results based on the two data sets are quantitatively close to each other. In this chapter I only show the results based on the later data set.

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 I 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 I turn to below after describing sample selection.

Sample Selection I 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. I 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. Some movies have to be left out because 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, my sample consists of 3,036 movies.

It is a good idea to include older movies as the "initial state" for the analyses. This is

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

[29] The web crawler is based on Scrapy, a open source framework for Python. For more information, visit http://scrapy.org.

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. I am 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 one dates back before the early 1980s. I have also tried using 1985-1994 as the initial period, but have not found qualitative changes to the results.

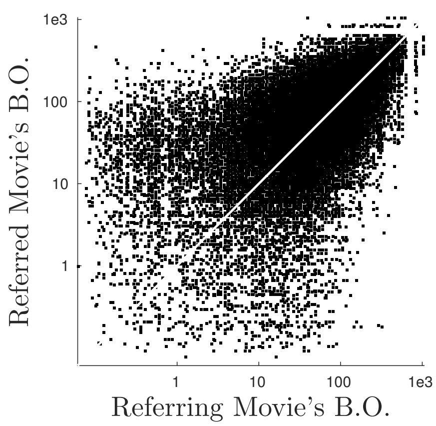
The Network Recall that the links between movies are constructed from references on IMDb and Amazon. Figure 2.1 provides a visualization of the references with regard to budget size. A dot represents a reference from one movie to another, where the box-office revenue of the first movie is given by the horizontal position of the dot, and the box-office revenue 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 Table 2.1, I show that the network captures similarity between movies in terms of the observed characteristics. The first two columns provide some descriptive statistics. For example, among the pairs of *linked* movies, 54.9% belong a same genre, in contrast to a 19.3% when it is among the pairs of any two movies. The last column presents a logit model that predicts whether a pair of movies are linked. All the coefficients are of expected signs and 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 .28. 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.

Table 2.1: Movie Pair Characteristics and Links

	All Dyads	Linked Dyads	Logit Model
Intercept			-5.51 (.035)
Same Production Company	3.60%	11.1%	0.624 (.048)
Same Rating	35.2%	62.4%	1.00 (.028)
Same Genre	19.3%	54.9%	1.36 (.028)
Same Leading Actor(s)	0.603%	29.5%	3.94 (.034)
Same Director(s)	0.099%	8.11%	3.12 (.09)
Same Writer(s)	0.081%	7.32%	2.99 (.09)
Diff. in Release Time (Year)	9.76	4.91	-0.146 (.003)
Diff. in Log Budget	1.22	0.724	-0.597 (.020)
Pseudo-R^2			0.275
N	9.63e7	3.03e4	9.63e7

In the cases that there are multiple production companies for one movie, the first-listed one is used. The same applies to genre. The last column is a logistic regression predicting linkages. Pseudo-R^2 equals 1 minus the ratio between residual deviance and null deviance. The entire sample (1975-2012) is included.



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

Figure 2.1: Budgets of the Referred Movies against Referring Movies

It is noted that the network is constructed from ex post data which studios did not possess when they green-lighted the movies. In respect to this, I do not assume that studios could use the ex post data. Instead, my assumption is that the studios understood the correlation structure represented by the network, and I use the ex post data here to back out that information. In fact, the constructed network 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.*[30] However, it does not link to, for example, *The English Patient* or *The Reader*, which also use WWII as background but lean toward a more romantic theme.

2.3 Reduced-form Analysis

In this section I present some model-free results. These results are not only interesting in their own right, but also motivate how the structural model will be specified in Section 2.4. At this point, I want to introduce some terminology that will prove useful later. 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, I examine 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 I take a look at the studio behaviors

[30] Both *Saving Private Ryan* and *The Shawshank Redemption* belong to the top guy-cry movies selected by *Entertainment Weekly*, 2005.

and explore what movies are more likely to imitate or be imitated by others.

In Table 2.2, Column 1 regresses the log ROI on a time trend, genre, MPAA rating, quality of the crew and log budget.[31] 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 one uses the log box-office revenue as the dependent variable, which is comparable to what was found in previous studies.[32]

In Column 2, I 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 Column 3, I keep the lag term but drop genre, rating and crew as covariates. Compared with Column 2, the decrease in R^2 is only marginal. Also notice that the coefficient for the lag is slightly increased. This is consistent with the observation that the network has incorporated, to certain extent, the proximity between movies with respect to the observed characteristics (Table 2.1). Partly for this reason, I will drop these covariates in some part of the structural model.

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 2.3 I regress the absolute value of the residuals from the last column of Table 2.2 on the number of precursors

[31] 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.

[32] 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 2.2: Spatial Regression of Log ROI

		(1)	(2)	(3)
Time	Constant	-0.694**	-0.544**	-0.551**
	Trend	0.0016	0.0085*	0.0111**
	Seasonality	0.136**	0.067*	0.075*
Log Budget		0.0787**	-0.0202	0.0116
Genre	...	Yes	Yes	
Rating	Restricted	-0.231**	-0.150**	
Crew	Actor	-0.0639	0.0501	
	Director	-0.0580	-0.0462	
	Writer	0.0921	-0.0390	
Average Log ROI of Precursors			0.821**	0.848**
R^2		0.056	0.257	0.247
N		2,943	2,943	2,943

** Significant at the 99% level. * Significant at the 90% 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 2.3: Polynomial Fit of Residual Size on Number of Precursors

	Absolute Residuals
Intercept	1.529 (.0723)
Number of Precursors	-0.195 (.0333)
Number of Precursors^2	0.0147 (.0048)
Number of Precursors^3	-4.95e-4 (2.5e-4)
Number of Precursors^4	6.03e-6 (4.4e-6)
Average Effect	-0.0770
N	2,943

Numbers in the parentheses are standard errors. The dependent variable is the absolute value of the residuals from the last column of Table 2.2. The average effect is the average derivative of the estimated polynomial across the movies in the data.

Table 2.4: Regression of the Number of Imitators / Precursors

		Log # of Precursors	Log # of Imitators
Time	Yearly Dummy	Yes	Yes
Genre	...	Yes	Yes
Rating	Restricted	0.0893**	0.0639**
Crew	Actor	0.174**	0.0233
	Director	0.0618*	0.0556*
	Writer	-0.0073	-0.0459
Log Budget		0.266**	0.232**
Log ROI			0.264**
R^2		0.439	0.525
N		4,390	4,390

** Significant at the 99% level. * Significant at the 90% level. See Table 2.2 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. 1 is added to the number of imitators or precursors before taking the log. The entire sample (1975-2012) is included.

(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 the 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 2.4 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

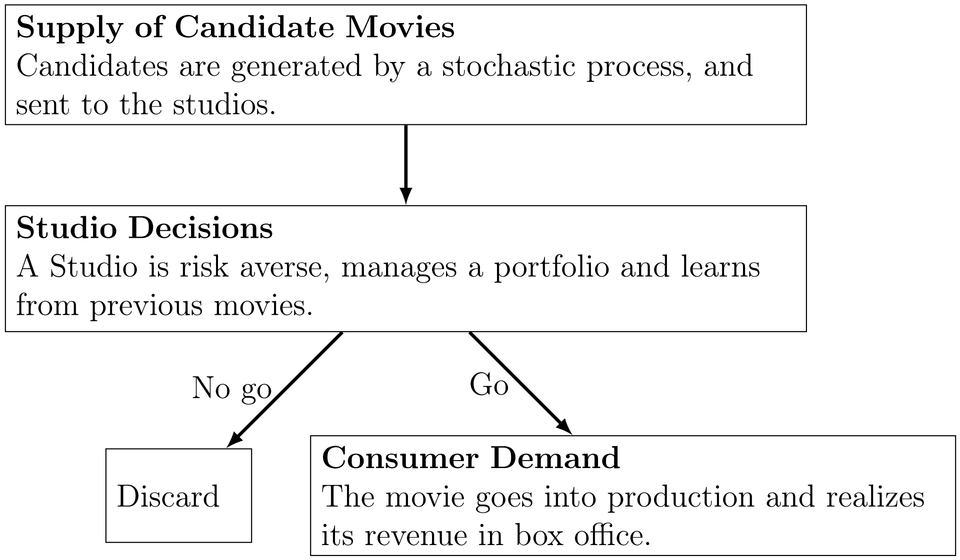


Figure 2.2: Model Overview

are more likely to be imitated, supporting the conventional wisdom that there is firm learning in the movie industry.

A more subtle point in Table 2.4 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, I will allow learning as well as risk aversion on the studio side.

2.4 Model

In this section I present an empirical model of product entry with learning. An overview of the model is given in Figure 2.2. 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. I 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.

2.4.1 Consumer Demand

The main objective here is to have a simple model for box-office performance. To start, let vector x_j collect the observed characteristics for movie j. Given my 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 ξ_j capture the average consumer taste for this short period over the characteristics that are not included in x_j. Let ε_{ij} be an idiosyncratic utility term. Consumer i's utility from movie j for this short period after j's release is

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

I explicitly model the correlation across the ξ's for different movies. The correlation structure is captured with the network, which I 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 ξ_j measures the excellence of the movie in the eye of consumers, I refer to it as the "latent quality."

The reduced-form analysis has shown 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 2.6. 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 I specify the utility as

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

Then, assuming type-I extreme value distribution for the idiosyncratic errors ε_{ij} and ε_{i0}, we have the market share of j given by $1/(1+e^{-U(x_j,\beta)-\xi_j})$. The box-office revenue equals the market share multiplied 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.[33] 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[34]

$$\pi_j = m_{r_j}/(1+e^{-U(x_j,\beta)-\xi_j}). \tag{2.4.1}$$

Note that there is a one-to-one relation between the box-office revenue π_j and latent quality ξ_j. So after a movie is exhibited in theaters, the box-office performance reveals the latent quality of the movie.

[33]See *Theatrical Market Statistics*, MPAA. I 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.

[34]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.

2.4.2 Candidate Products

Instead of pre-fixing a product space where firms can choose from, I 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.

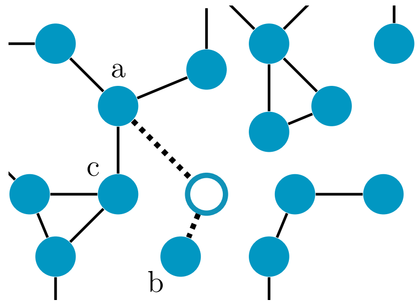
Treating time as continuous, I let candidate movies arrive at a Poisson rate η_f for firm f.[35] Suppose that a candidate movie j arrives at time t for firm f. If it ever gets produced, its arrival time is recorded as a_j and its production period is $(a_j, r_j]$. The observed characteristics x_j, latent quality ξ_j, as well as the location of j in the similarity network, are drawn from a state-dependent distribution. The state is denoted as $\mathscr{S}_t$ and is the collection of the observed characteristics, latent qualities and network for all the existing movies at t (released or in-production).

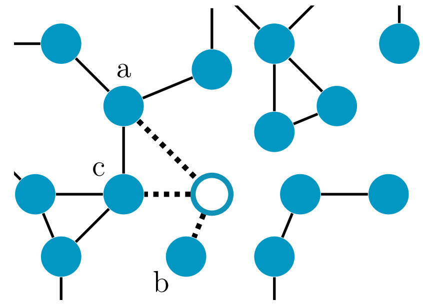
I first draw the candidate's location in the network, namely what existing movies are similar to j. Then, the characteristics x_j and ξ_j are drawn conditional on the location. In this sense, each candidate is either completely novel or a "creative combination" of some existing products.[36] In principle, 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, I want to point out that the arrival process is a latent structure. I have experimented with many variations of the process and my choice has been guided by both economic intuition

[35] The model allows a different arrival rate for each firm, which introduces quite a lot parameters. For estimation, I 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.

[36] 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).





Solid nodes represent existing movies. The arrival, represented by a hollow node, attaches to each existing movie independently in the first stage. A realization is displayed on the left where it attaches to node (a) and (b). Given this outcome, the arrival further attaches to each neighbor of (a) in the second stage. A realization is displayed on the right where it attaches to (c).

Figure 2.3: An Illustration of the Two-Stage Process

and patterns in the data. The extent to which the model is capable of reproducing the data is partially assessed in Section 2.6.2.

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, I 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. I let the attachment probability follow a logit model:

$$\Pr\left(y_{k,j} = 1 | \mathscr{S}_t\right) = \frac{1}{1 + e^{-F(x_k, t; \gamma)}}, \quad \forall k. \tag{2.4.2}$$

I 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 were previously shown in Table 2.1.

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 my data. The average clustering coefficient[37] is 0.24. 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.

I 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 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 ω. An example of the two-stage process is illustrated in Figure 2.3. In the Appendix, I show how to calculate the 1st-stage probabilities from (2.4.2).

Observed Characteristics Given the network location for j, now I specify a distribution $\Pr(x_j|y_j,\mathscr{S}_t)$ 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

[37] 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).

a group of big-budget movies should likely be a big-budget. Had genre been included in x_j on the supply side, the same logic applies to a candidate whose precursors are all sci-fi or comedy movies, for example.

Specifically, the budget b_j is drawn from a truncated normal distribution. The variance-to-mean ratio, denoted by χ, 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 μ, which is another parameter to be estimated.[38] The release time is determined by $r_j = a_j + \Upsilon(b_j)$, where Υ is a function that relates production time with budget size, and is non-parametrically estimated "off-line" with the data on production start date. This is a simplification and I abstract away from the factors that influence the release date after the green-light decision.

Latent Quality Now I specify a distribution $\Pr(\xi_j|x_j, y_j, \mathscr{S}_t)$ from which the latent quality of j is drawn. Here one wants to consider two factors. First, ξ_j should be correlated with ξ_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. So to the extent that consumer tastes can be changing over time, the difference in the release time, $|r_j - r_k|$, should be able to dilute the correlation between ξ_j and ξ_k. It seems appropriate to make the following specification:

$$\Pr(\xi_j|x_j, y_j, \mathscr{S}_t) = \mathscr{N}\left(\frac{\lambda \sum_{y_{k,j}=1} \varphi^{|r_k - r_j|}\xi_k}{1 + \lambda \sum_{y_{k,j}=1} \varphi^{|r_k - r_j|}}, \frac{\sigma^2}{1 + \lambda \sum_{y_{k,j}=1} \varphi^{|r_k - r_j|}}\right), \tag{2.4.3}$$

Note that (2.4.3) resembles the Bayesian updating formula under normality, where the latent quality of a precursor, ξ_k, is treated as a signal for ξ_j. The weight for each signal is calibrated by λ, which can be thought of as a measure of correlation.

[38]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. I 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.

For the extreme case $\lambda = 0$, the ξ'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.

2.4.3 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.[39] My corresponding assumption is that firms do not observe the latent quality for either the candidate or any movie that is still in production. However, I 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, let us start with the information set for the firms, denoted as $\mathscr{F}_t$. 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. $\mathscr{F}_t$ also includes the box-office revenue π_k, or equivalently, the latent quality ξ_k of each movie k that has been released. Notice the important difference between $\mathscr{S}_t$ and $\mathscr{F}_t$ that the later does not include the latent qualities of the movies that are still in production.

Given the information set, one can work out the belief for the firms. For the simplest case where one is looking at a single candidate j whose precursors have all been released, the belief $\Pr(\xi_j|\mathscr{F}_t)$ is simply given by (2.4.3). 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

[39] 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)

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: $\Pr(\{\xi_k : a_k \leq t, r_k \geq t\}|\mathscr{F}_t)$. A nice feature of my model is that there is a closed-form expression for this joint density, which I derive in the Appendix.

Given the belief, now one can model the firm's investment decisions. Given $\mathscr{F}_t$, denote by $P_{f,t} \equiv \{k \in f : a_k < t, r_k \geq t\}$ the set of f's in-production movies, which can be thought of as f's portfolio. The firm's problem is whether it is desirable to add j into the portfolio. The present value of $P_{f,t}$ is

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

where δ is a discounting factor and π_k is given by equation (2.4.1). 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 π's depend on the ξ's, which the firms does not know. Let b_j be the production budget of j, and ζ_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

$$\mathbb{E}\left[V\left(\Pi(P_{f,t} \cup \{j\}) - b_j - \zeta_j\right)\middle|\mathscr{F}_t\right] > \mathbb{E}\left[V(\Pi(P_{f,t}))\middle|\mathscr{F}_t\right], \tag{2.4.4}$$

where the expectation is taken over the ξ's. I specify V to bear a constant coefficient of absolute risk aversion (CARA), denoted as α. I 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 (2.4.4) 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 essay. I leave it for future research.

One can readily define a risk-free equivalence of the revenue π_j, denoted by $\overline{\pi}_j$, through the equation

$$\mathbb{E}\left[V(\Pi(P_{f,t})+\overline{\pi}_j)\middle|\mathscr{F}_t\right]=\mathbb{E}\left[V(\Pi(P_{f,t}\cup\{j\}))\middle|\mathscr{F}_t\right].$$

When $P_{f,t}$ is empty, it reduces to the more familiar equation: $\mathbb{E}\big(V(\pi_j)|\mathscr{F}_t\big)=V(\overline{\pi}_j)$. The definition allows us to state condition (2.4.4) alternatively as $\overline{\pi}_j-b_j-\zeta_j>0$. We may also define $\overline{\pi}_j/b_j$ 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 $\overline{\pi}_j/b_j=1$ 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 one halves the acceptance probability for every candidate and doubles the arrival rate at the same time. Effectively, one can only identify the *relative* acceptance probabilities for different movies. In the Appendix, I use a simplified version of the model to further explain why the normalization is innocuous and necessary.

2.5 Model Estimation

2.5.1 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 ξ by arguing that firms have no knowledge of it before entry (see, for example, Aguirregabiria and Ho, 2011; Eizenberg, 2014). Because my model relaxes this assumption, it requires an extension of the standard estimation technique.[40]

To be more specific, notice that the following regression equation can be directly obtained from the box-office equation (2.4.1).

$$\log\left(\pi_j\right) - \log\left(m_{r_j} - \pi_j\right) = U(x_j, \beta) + \xi_j. \tag{2.5.1}$$

There are two issues about this regression. First, the residuals ξ are correlated, and the correlation structure is of interest to us. Second, due to endogenous entry, the standard moment condition $\mathbb{E}(\xi_j | x_j) = 0$ does not hold here in general. For example, ξ_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 ξ_j for entry.

I estimate this equation by controlling for what firms can learn about ξ_j at the time of entry. In my model, the firm's information at time a_j is a subset of $(x_j, y_j, \mathscr{S}_{a_j})$, which determines the arrival distribution of of ξ_j through (2.4.3). Let $v_j \equiv \xi_j - \mathbb{E}(\xi_j | x_j, y_j, \mathscr{S}_{a_j}; \beta, \lambda, \varphi)$, which is the difference between the realized latent quality and the mean of its arrival distribution. I use v instead of ξ to construct moment conditions. Identification of parameter σ requires us to further match the dispersion of the

[40] 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).

arrival distribution, so I define a second difference: $\iota_j \equiv \upsilon_j^2 - \mathbb{E}(\upsilon_j^2 | x_j, y_j, \mathscr{S}_{a_j}; \beta, \lambda, \varphi, \sigma)$. The demand-side estimation is then based on the following mean-independence moments:

$$\mathbb{E}\Big((\upsilon_j, \iota_j) \,|\, x_j, y_j, \mathscr{S}_{a_j}\Big) = 0.$$

This leaves us with many instruments to choose from the conditioning set to interact with υ_j and ι_j.[41] To identify β, I interact υ_j with x_j. To identify parameter λ and φ, I interact υ_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 σ, I interact ι_j with a constant term and the number of the precursors for j. To the extend that the ξ's are unobserved, they need to be computed through equation (2.5.1) as a function of the unknown parameter β. This means that the estimation requires a numerical search jointly over $(\beta, \lambda, \varphi, \sigma)$. However, lots of computational time can be saved by using the OLS estimates of (2.5.1) as the initial guess for β.

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.

2.5.2 Supply

The estimation procedure for the supply side is relatively straightforward. I 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 $(x_j, y_j, \xi_j, \mathscr{S}_{a_j})$. Let H be a function this history. For notation, write $H(x_j, y_j, \xi_j, \mathscr{S}_{a_j})$ as H_j. The specification of H depends on the moments that one wants to match to identify the parameters. I give the specification of H below after discussing the

[41] For reasons why one should not use many moment conditions, see Andersen and Sørensen (1996) and more recently Han and Phillips (2006).

identification.

Collect the supply-side parameters in Θ. For any value of Θ, given the history at time a_{j-1}, the model makes a prediction of H_j, where the error of prediction given by

$$h(x_j, y_j, \xi_j, \mathscr{S}_{a_j}; \Theta) \equiv H_j - \mathbb{E}(H_j | x_{j-1}, y_{j-1}, \xi_{j-1}, \mathscr{S}_{a_{j-1}}; \Theta).$$

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.[42] The supply-side estimation relies on the following moment conditions:

$$\mathbb{E}\Big(h(x_j, y_j, \xi_j, \mathscr{S}_{a_j}; \Theta)\Big) = 0.$$

The estimate for Θ 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 my 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 I 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, I show through Monte Carlo experiments that the parameter values can be recovered from simulated data.

For the arrival process, the second-stage attachment probability (ω) 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 ω. The identification of the arrival rate (η) relies on the normalization that

[42] 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. I chose to use 100 simulations. Nevertheless, one evaluation of the objective function takes around 10 mins on a quad-core desktop.

a movie with $\overline{\pi}_j = b_j$ is accepted with .5 probability. As already pointed out, the model is observationally equivalent if one halves the acceptance probability for each candidate and doubles the arrival rate at the same time. In this regard, one should not interpret the estimated η literally as the number of the proposals that are sent to studios over a year.

For the selection model, the coefficient of risk aversion (α) 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 my 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 (ρ). 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 ρ decreases.

The specification of H complies with my identification argument. Specifically, I 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.

2.6 Estimation Results

2.6.1 Parameter Values

Table 2.5: Model Parameter Estimates, Demand-Side

Parameters		I	II	III
Time	Constant	-7.14 (.14)	-7.50 (.18)	-7.51 (.17)
	Trend	-0.0202 (.004)	-0.0185 (.008)	-0.0156 (.008)
	Seasonality	0.151 (.05)	0.0941 (.04)	0.0962 (.04)
Budget		Yes (Fig. 2.4)	Yes (Fig. 2.4)	Yes
Rating	Restricted	-0.197 (.05)	-0.231 (.06)	
Genre	Drama	0	0	
	Comedy	0.418 (.07)	0.248 (.08)	
	Action/War	0.197 (.07)	0.254 (.08)	
	Family	0.386 (.09)	0.596 (.16)	
	Sci-Fi/Advent.	0.358 (.1)	0.358 (.09)	
	Horror	1.11 (.1)	0.914 (.2)	
	History/Bio.	0.106 (.1)	0.0367 (.1)	
	Others	-0.424 (.4)	-0.405 (.2)	
Crew Power	Actor	-0.0266 (.05)	0.012 (.05)	
	Director	-0.0679 (.06)	-0.0667 (.05)	
	Writer	0.121 (.06)	0.0187 (.05)	
Similarity (λ)			0.571 (.07)	0.624 (.07)
Yearly Disc. Factor (φ)			0.963 (.02)	0.954 (.02)
Std. Dev. (σ)			1.86 (.05)	1.91 (.05)
R^2 (Share)		0.557	0.681	0.676
R^2 (ROI)		0.0716	0.331	0.321

Column I displays the OLS estimates of equation (2.5.1). Column II and III display the GMM estimates. See Table 2.2 for definitions of some of the variables. The utility for budget is estimated as a piecewise linear function; see Figure 2.4 for the estimates (dashed curve for Column I and solid curve for Column I). 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.

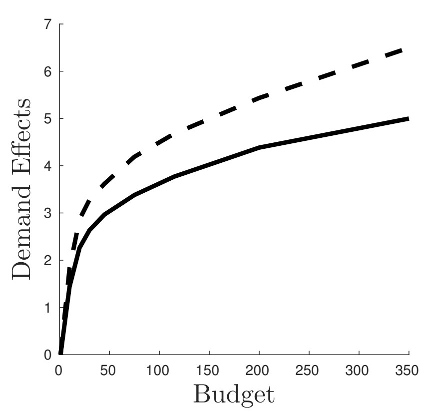


Figure 2.4: Estimated Effects (β) for Budget Size

Demand side Table 2.5 displays the estimates for the demand-side parameters. Specifically, Column I displays the estimates from an OLS regression of the revenue equation (2.5.1); 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 rating reduces demand. Interestingly, horror movies are the best bet for studios to make profits.[43] The effect of star power is insignificant, which is consistent with my 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 2.4. The shapes of the function exhibit diminishing marginal utility.

The difference between the GMM and OLS estimates for the effects of budget (Figure 2.4) can be explained by studio selection. For example, for a big-budget movie to be produced, a high belief on ξ is typically required to compensate the associated large risks. This introduces a positive correlation between b_j and ξ_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 2.5). 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 explanator 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

[43]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*.

Column II provides a considerable improvement. This is in line with my reduced-form results (Table 2.2). As we move from Column II to III, both the R^2's remain almost identical, and λ picks up the effects of the dropped covariates. This is again in line with my 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 σ implies an enormous level of uncertainty. To see the magnitude, recall that σ 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 λ and φ, the model implies that for the movies in the data, learning on average reduces the variance in ξ at the time of go/no go decision by about 60%.[44] The estimate of φ 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 2.6 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 γ_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.

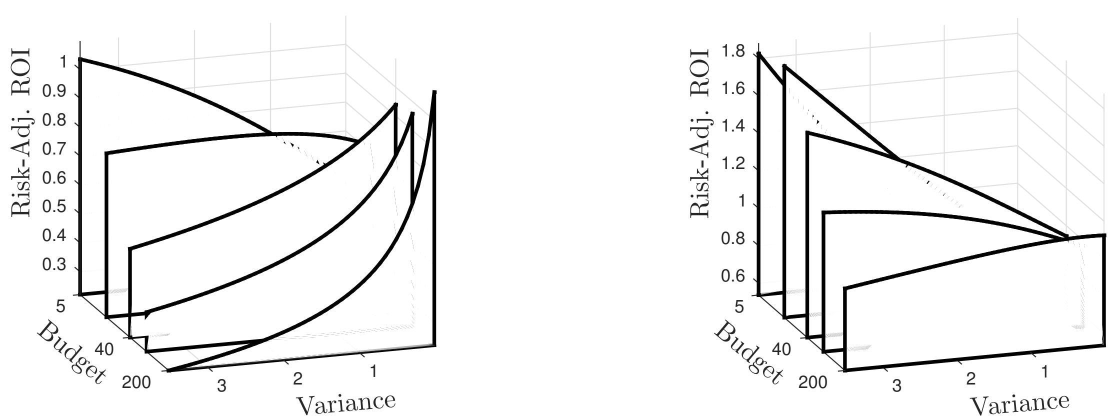
The coefficient of absolute risk aversion, α, is estimated to be both statistically and economically significant. This is demonstrated in Figure 2.5. On the left side, the graph plots the risk-adjusted ROI as a mapping of the budget size and variance in the ξ. We see that for big-budget movies, risk-adjusted ROI decreases with the variance in ξ. This should be expected because a higher variance implies a higher level of

[44] More precisely, for each movie, I compare $\mathbb{V}\mathrm{ar}(\xi_j|\mathscr{F}_{a_j})$ with σ^2, which is the variance that the firm would perceive had it ignored the correlation structure in ξ.

Table 2.6: Supply-Side Parameter Estimates

Parameters		Estimates	
Attachment	2nd-Stage Probability (ω)	0.215	(.005)
	Intercept (γ_0)	-4.836	(.02)
	Own Movies (γ_1)	1.673	(.06)
	Time Difference in Years (γ_2)	-0.214	(.005)
Obs. Characteristics	Budget Mean w/o Precursors (μ)	60.6	(8.5)
	Variance-to-Mean Ratio (χ)	103.1	(6.9)
Coefficient of Risk Aversion (α)		0.0312	(.0051)
Standard Deviation of Decision Shock (ρ)		0.601	(.03)
Yearly Arrival Rate (η)		772	(45)

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 δ is set at .975. Numbers in the parentheses are standard errors computed by parametric bootstrapping (see Appendix).



The plots show the risk-adjusted ROI, $\overline{\pi}_j / b_j$, of a hypothetical movie j as a function of the budget size and variance in the latent quality ξ_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 2.5: Risk-adjusted ROI as Function of Budget Size and Var(μ)

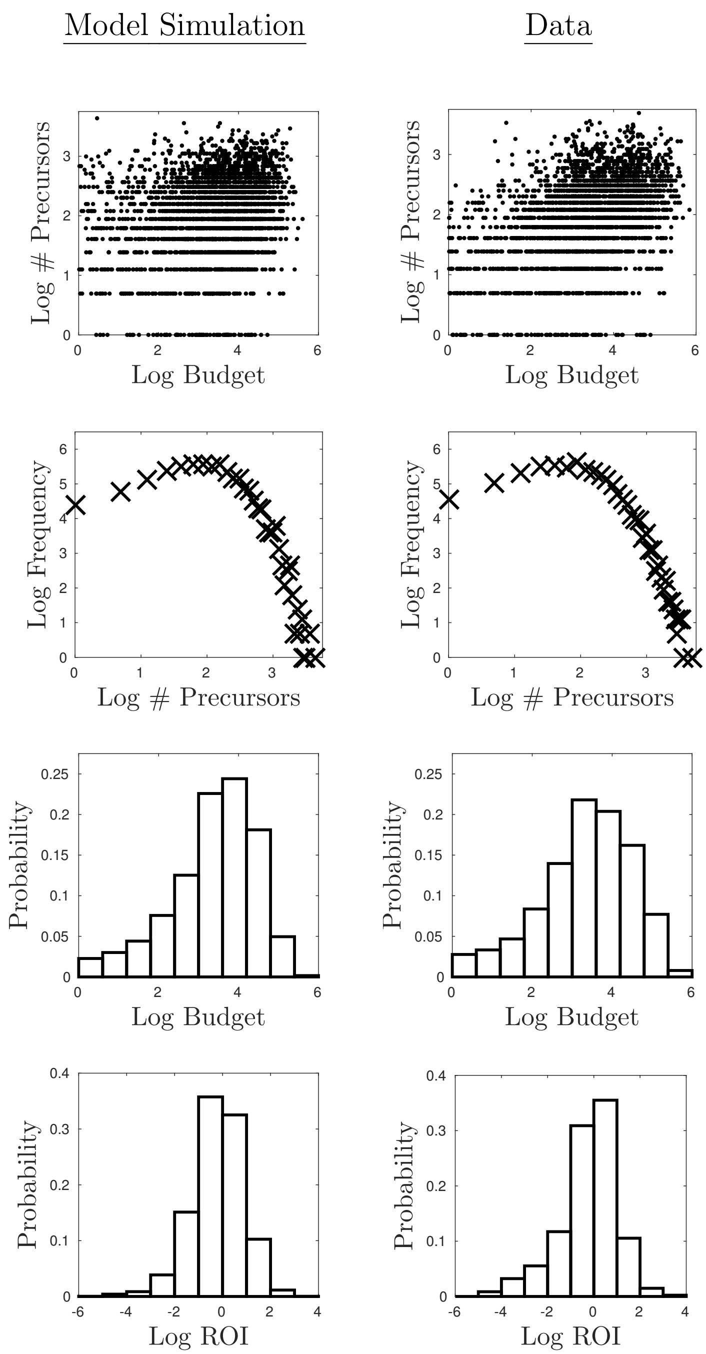
risks. However, what may come at a surprise is that this relation is reversed for small-budget movies. This is because there is a tight lower bound for how much a small-budget movie can lose: the maximum that a movie with \$5 million budget can lose is \$5 million. In this case, a larger variance expands 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.[45]

As a comparison, the graph on the right side in Figure 2.5 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 I 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, ρ, the less predictive is the model about the firm decisions. In the extreme case where $\rho \to +\infty$, all types of arrivals are accepted with equal probability. Under the estimate of ρ, the acceptance probability ranges from about 0 to .55 for the range of the risk-adjusted ROIs shown in the left graph of Figure 2.5. So my 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.[46]

[45] 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.

[46] 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 $a_k < a_j$. 1 is added to the number of precursors before taking its log.

Figure 2.6: Model Fit

2.6.2 Model Fit

To make an assessment on model fit, I simulate the model from 1995 all the way to 2012, conditional on the data from 1975 to 1994. Figure 2.6 compares the simulated data with the real data. Given the important trade-off between budget size and uncertainty in the firm's decision (see Figure 2.5), here let us focus on: (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 the supply model, I find the fit satisfactory. Because the production 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 my 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 ξ. 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.

2.7 Counter-factual Experiments

This section uses several counter-factual experiments to provide further insights on how learning affects product entry. First, I quantify the importance of learning in the model by investigating what happens if firms stop learning. I also try to quantify the spillover effects of learning. Second, I look at what happens if there is a change in firm's risk attitude. The results provides explanation to the rise of imitative block-busters in the last decades.

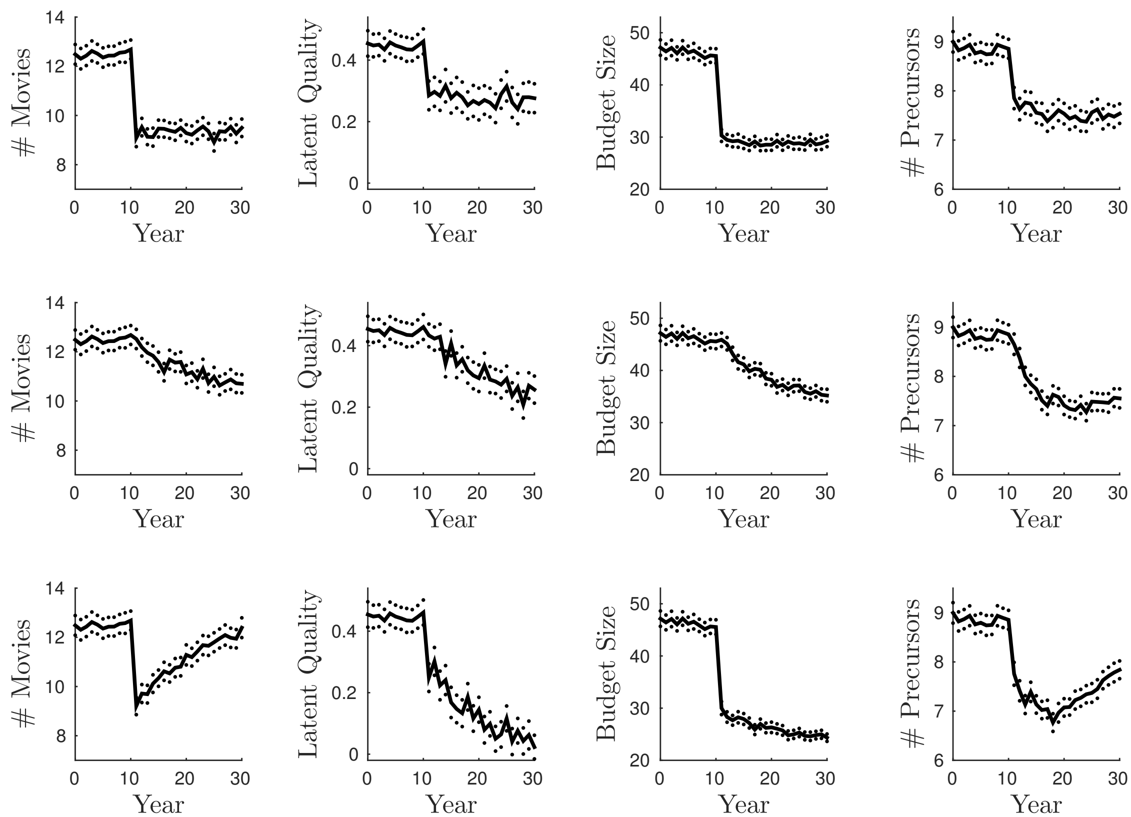
For all the counter-factuals, the changes are introduced at the steady state of the model. For the model to have a steady state, the demand trend is removed and both the market size and ticket price are set constant over time. The rest parameters are set at their estimates. To reach the steady state,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

2.7.1 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? The first set of counterfactuals are designed to answer these questions. I first examine the case where a single firm stops learning, which is illustrated in the top row of Figure 2.7. 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 ξ, 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 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 2.7. 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



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 2.7: What Happens If Firms Stop Learning

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 even though it is estimated that there are significant barriers against social learning, in the sense that a studio has a favorable position in imitating its own movies.[47]

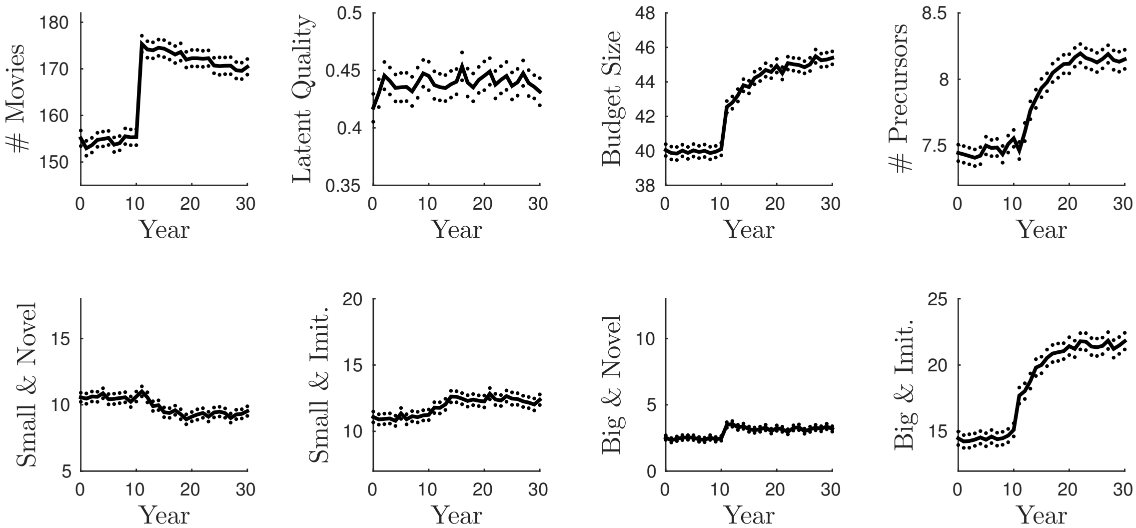
Finally, the bottom row of Figure 2.7 displays the case where all the firms stop learning. As expected, we see much larger decreases in budget size and latent quality. Notice that after the initial drop, there is a gradual *increase* in the number of movies produced. This is caused by the larger decrease in the budget size, which, combined with the mis-perceived novelty in the candidates, makes the firms see high profitability (Figure 2.5) and accept more candidates.

2.7.2 Risk Aversion

Given the important role of risk aversion in the industry as well as the model, I 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 2.8 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.

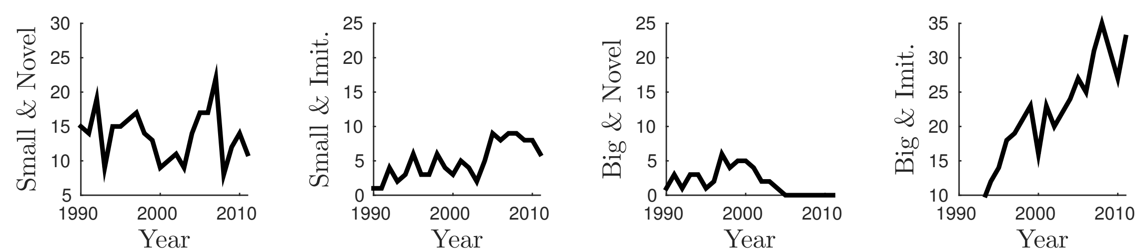
As expected, there are increases in both the number and budget size of the movies produced each year: as the firms become less risk averse, they accept more candidates. In particular, they become more accepting towards bigger-budget candidates, as they involve higher level of risks than their low-budget counterparts. The releases of these

[47] Though in this essay the focus is on 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 2.7. 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 2.8: 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 2.7.

Figure 2.9: Rise of Big-Budget Imitative Movies

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 this rise of imitativeness, I want to draw some 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 2.5). 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 2.8. 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 2.9 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, we 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 the 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."[48] Spielberg and Lucas, among others, expressed concerns over the declining originality in motion pictures, and point the finger at risk-focused studios.[49] However, my 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.[50] 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 2.9). 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.

2.8 Concluding Remarks

This essay studies product entry in the presence of firm learning from the demand for previous products. I 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, I am able to measure the imitativeness of different products. I demonstrate the use of evolving networks in modeling industry dynamics. The model, together with its estimates, allows me to quantify the effects of learning, generate insights into

[48] "Are Blockbusters Destroying the Movies." *New York Times*, Jan 6, 2015.

[49] See "George Lucas & Steven Spielberg: Studios Will Implode; VOD Is the Future," *Variety*, June 12, 2013; "Steven Spielberg and George Lucas Predict Film Industry Implosion," *The Guardian*, June 13, 2013.

[50] Co-financing is not explicitly modeled in this essay. Interested readers may look at Goettler and Leslie (2005).

the balance of innovation vs. imitation, and examine the role of risk aversion. Given my focus on the U.S. motion picture industry, it is natural to ask how my 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 the essay's insight on the balance between innovation and imitation, Alberts (2012) argue that 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. 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).

Appendix A

Appendix to Chapter 1

A.1 Alternative Models

Let us first consider a two stage model, where carriers set frequencies in the 1st stage and compete on prices in the 2nd stage. In such setting, frequency must be an explicit choice variable. One can easily transform it into a simultaneous game where carriers set frequencies and prices together. However, it should be noted that it is strategically different from the two stage game, which could give rise to preemptive behaviors, for example.

Fix the network structure. Let p_j be the price on route j, and F_g be the flow capacity on link g (which we can think of as flight frequency). Both route and link are carrier-specific. Use $j \in c$ to denote that route j belongs to carrier c and $g \in j$ to denote that link g is a segment of j. As in BLP, we may allow a route-specific demand component ξ_j. Let $D_j(p,\xi,F)$ be the demand for route j. There are costs associated with setting up capacity that are sunk in the second stage. Let ω_j be a route-specific marginal cost in addition to the capacity costs. Each carrier's problem in the second stage is then

$$\max_{\{p_j, j\in c\}} : \sum_{j\in c} D_j(p,\xi,F)\cdot(p_j-\omega_j)$$

$$\text{s.t.} \sum_j D_j(p,\xi,F) \cdot \mathbf{1}\{g \in j\} \leq F_g, \text{ for each } g \in c.$$

Let $\Pi_c(\xi, F)$ be the second-stage profit at the equilibrium prices for carrier c. Next consider the first stage. Let C_g be the marginal cost of capacity on link g, each carrier's problem is then

$$\max_{\{F_g, g\in c\}} : \Pi_c(\xi,F) - \sum_{g\in c} C_g \cdot F_g.$$

The main challenge of this formulation are the inequality constraints in the second stage. To characterize an equilibrium in price, one needs to rely on Kuhn-Tucker conditions instead of the simpler first order conditions. Further, for the first stage, the profit function $\Pi_c(F)$ is not differentiable at the places where the constraints are exactly binding (or exactly non-binding). As a result, it does not seem straightforward to extend the BLP framework to such a formulation.

In a simultaneous game, each carrier's problem becomes the following. Still, it seems that one needs to rely on Kuhn-Tucker conditions to characterize an equilibrium.

$$\max_{\{p_j, j\in c\}, \{F_g, g\in c\}} : \sum_{j\in c} D_j(p,\xi,F) \cdot (p_j - \omega_j) - \sum_{g\in c} C_g F_g$$

$$\text{s.t.} \sum_j D_j(p,\xi,F) \cdot \mathbf{1}\{g \in j\} \leq F_g, \text{ for each } g \in c.$$

Let us consider the assumption that the carriers always choose F so that the constraints are exactly binding. Intuitively, this should be the case when capacity costs C_g's are not so small that carriers find empty flights desirable, nor so large that carriers have to reject some passengers with overly high prices. If we are willing to make this assumption, then the carrier's problem is reduced to:

$$\max_{\{p_j, j\in c\}} : \sum_{j\in c} D_j(p,\xi,F) \cdot \left(p_j - \sum_{g\in j} C_g - \omega_j\right)$$

$$\text{where } F_g = \sum_j D_j(p,\xi,F) \cdot \mathbf{1}\{g \in j\}, \text{ for all } g.$$

Notice that this essentially imposes a fixed point requirement on D and is exactly the model that I have considered in the essay.

A.2 Uniqueness of The Fixed Point

Here I consider the fixed points for mapping $\Psi(p,\xi,\cdot)$ over the positive demand vectors. I first present a condition on the parameters which guarantees a unique fixed point regardless of the network configuration.

Proposition 1. *If* $\max\{\theta_1,\theta_2\} < \frac{\lambda}{2}$*, then* $\Psi(p,\xi,\cdot)$ *has a unique fixed point. Moreover, Iteration of* $\Psi(p,\xi,\cdot)$ *always converges to the fixed point.*

The parameter estimates are $\theta_1 = 0.33$, $\theta_2 = 0.49$ and $\lambda = 0.56$, so the condition is not satisfied. However. it is a sufficient condition and the proof considers the worst case scenarios that could break convergence to a unique fixed point. For example, if one is willing to assume $S \ll 1$ so that $D_j \simeq M_m \cdot S^{\lambda-1} e^{v_j/\lambda}$ is a good approximation for (1.2.5), then $\max\{\theta_1,\theta_2\} < \frac{\lambda}{2-\lambda}$ is sufficient to obtain the results in the proposition. Alternatively, if one considers networks where there is a single route in each market, then $\max\{\theta_1,\theta_2\} < 1$ is sufficient.

In my experiences, under the estimated parameter values, iteration of $\Psi(p,\xi,\cdot)$ always converges to the same point regardless of the starting point. For the starting point, I have tried the observed demand vector, a vector of ones, and many random vectors. When the values of θ_1 and θ_2 are much larger, for example, $\theta_1,\theta_2 > \lambda$, the iteration typically diverges or converges to non-positive vectors.

Below I give the proof of the proposition.

Proof of Proposition 1. First let us define a "distance" measure between two positive vectors of the same length n:

$$d(V,V^*) := \max\left\{\frac{V_1}{V_1^*},\dots,\frac{V_n}{V_n^*},\frac{V_1^*}{V_1},\dots,\frac{V_n^*}{V_n}\right\}.$$

Note that d is always greater than or equal to 1, and $d = 1$ implies $V = V^*$.

Let D and D^* be two positive demand vectors. With the specification in (1.2.1), it is seen that

$$d\left(F(D), F(D^*)\right) \le d(D, D^*).$$

With the specifications given by (1.2.3) and (1.2.4), it is seen that

$$d\left(e^{f(D;\theta,\sigma)}, e^{f(D^*;\theta,\sigma)}\right) \le d\left(F(D), F(D^*)\right)^{\max\{\theta_1,\theta_2\}},$$

where the exponential e^f refers to the vector collecting e^{f_j}, $j = 1, ..., n$. Using the same notation for e^v, we have

$$d\left(e^{v(p,\xi,D)}, e^{v(p,\xi,D^*)}\right) = d\left(e^{f(D;\theta,\sigma)}, e^{f(D^*;\theta,\sigma)}\right).$$

It is not immediate but neither difficult to see that the nested logit specification implies

$$d\left(\Psi(p,\xi,D), \Psi(p,\xi,D^*)\right) \le d\left(e^{v(p,\xi,D)}, e^{v(p,\xi,D^*)}\right)^{2/\lambda}.$$

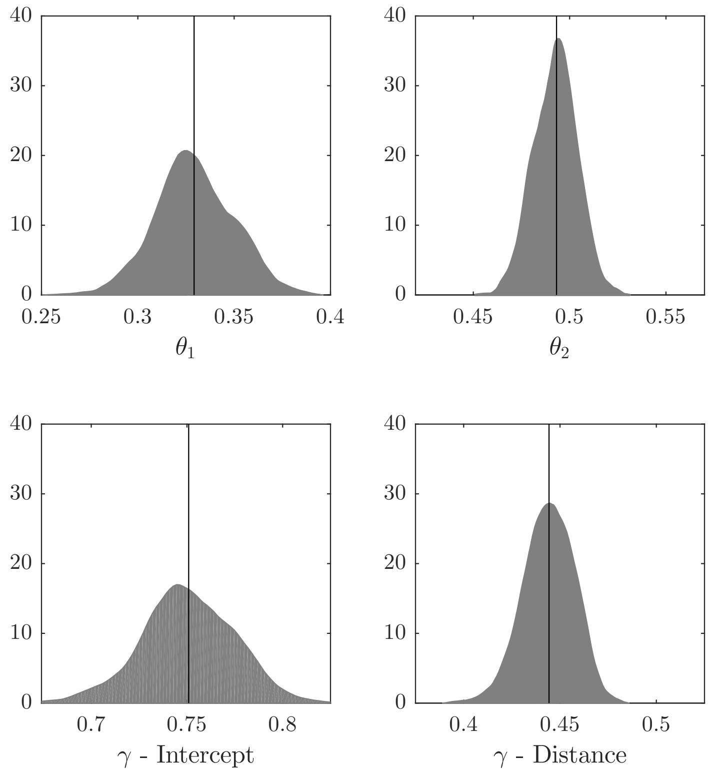
Putting this chain of inequalities together, it is seen that under the condition of the proposition,

$$d\left(\Psi(p,\xi,D), \Psi(p,\xi,D^*)\right) < d(D, D^*).$$

Because Ψ shrinks the distance between D and D^*, it cannot have two distinct fixed points. □

A.3 Monte Carlo Experiment

The main purpose of the Monte Carlo experiment is to check whether the estimator becomes unworkable or ill-behaved with the network structure and fixed-point demand. I use the benchmark estimates as the "true" parameter values for the experiment, so the results can also be used to construct bootstrap standard errors.



Notes: The experiment is repeated for 500 times. Each plot shows the kernel density (with optimal bandwidth) of the estimator for a particular parameter. The vertical line is the “true” parameter value.

Figure A.1: Monte Carlo Results

More specifically, the experiment is done in the following manner. A network structure and a set of "true" parameter values are chosen before the experiment. To start the experiment, a set of ξ_j's and ω_j's is drawn. Then an equilibrium in price is computed (see Section 1.5.1). Treating the equilibrium demand and price as data, one can recover the parameter values using the estimation procedure outlined in Section 1.3. This samples one observation from the distribution of the estimator. Repeat the experiment with another set of ξ_j's and ω_j's to sample another observation.

For a large network such as the one used in the estimation, computation an equilibrium is very intensive. It would take me too much time to obtain a sample of reasonable size for the estimator. For this reason, I conducted the experiment with a smaller network. Specifically, I use a more stringent definition where a link is included iff its flow is larger than 2.5 percent of the geometric mean of the population of the end cities. The resulting "backbone" network still includes the 12 carriers, but only 602 links and $n^b = 7{,}353$ routes. For a network of this size, it is possible to compute $\partial D/\partial p$ with implicit function theorem (see Section 1.3.1), which saves a lot of time.

Because n^b is smaller than n, the number of routes in the data, I downsize the standard deviations of ξ_j and ω_j by a factor of $\sqrt{n^b/n}$. In generating the ξ_j's and ω_j's, one possibility is to draw them independently. However, one can imagine that there are correlations between the ξ_j's within the same market, and between the ω_j's that share a common link. Such correlations are indeed present in the data. Accordingly, I introduce these correlations in the generated shocks by using market-specific and link-specific dummies.

Figure A.1 shows the kernel densities of the estimator for four parameters: θ_1, θ_2, and two cost side coefficients in γ. The densities are all bell shaped and centered around the "true" parameter values. The variance of the estimator for θ_2 is smaller than that for θ_1 because there are more one-stop routes than non-stop routes. Also note that variances of the estimator for γ partly come from the demand-side estimation. Overall, the estimator seems well-behaved.

Appendix B

Appendix to Chapter 2

B.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 η. 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, ξ_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 ξ_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. Assume that $z_j \sim \mathcal{N}(0,\rho^2)$. The firm accepts j iff $\bar{\xi}_j - z_j > 0$, and discards j otherwise.

The model has five parameters: η, n, λ, σ and ρ. The question is whether one 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 $v_j \equiv \xi_j - \lambda\xi_{y(j)}$

is zero-mean normal with variance σ^2 and is i.i.d. across the accepted j, one can identify λ and σ by simply regressing ξ_j on $\xi_{y(j)}$ for $j \in A$. Next, because a smaller ρ makes the firm more selective in accepting products, the average expected log return of the accepted product, $\frac{1}{\#A}\sum_{j\in A}\overline{\xi}_j$, can be used to identify ρ. In the extreme case $\rho = +\infty$, there is no selection and the average should converges to 0. Finally, given n, λ, σ and ρ, the production rate $\#A/T$ is strictly increasing in the arrival rate so it can be used to identify η.

When there are additional revenues or costs that are proportional to the ones used in calculating ξ, one can model them by adding an intercept parameter to the firm's decision. A product is accepted iff $\overline{\xi}_j - z_j - c > 0$. The question here is whether c can be identified. From econometrician's perspective, the acceptance probability is

$$\begin{aligned}\Pr(j \text{ is accepted}) &= \Psi\left(\frac{\overline{\xi}_j - c}{\rho}\right) \\ &\simeq \Psi\left(-c/\rho\right) + \frac{\psi\left(-c/\rho\right)}{\rho}\cdot\overline{\xi}_j,\end{aligned}$$

where Ψ (ψ) is the cdf (pdf) of the standard normal distribution. The second line is a linear approximation of the probit model around $\overline{\xi}_j = 0$. 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:

$$\begin{aligned}\Pr(j \text{ is accepted})' &\simeq \Psi(0) + \frac{\psi(0)}{\rho'}\cdot\overline{\xi}_j \\ &\simeq \frac{\Psi(0)}{\Psi(-c/\rho)}\times\Pr(j \text{ is accepted}).\end{aligned}$$

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

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

then the two sets of parameters are almost observationally equivalent. Had a linear probability model been specified in place of the probit model, parameter c would not be identified.

B.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

$$1 - \Pr(y_{k,j} = 1 | \mathscr{F}_t) = (1 - p_{k,j}) \cdot \prod_{\ell \sim k} (1 - p_{\ell,j}\omega),$$

where $\ell \sim k$ indicates that ℓ and k are linked in Y.

In principle, given the value of $\Pr(y_{k,j} = 1 | \mathscr{F}_t)$ for all k (as specified by (2.4.2)), 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,

$$1 - \Pr(y_{k,j} = 1 | \mathscr{F}_t) \simeq (1 - p_{k,j})(1 - p_{k,j}\omega)^{d_k(Y)},$$

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

$$\log\left[1-\Pr(y_{k,j}=1|\mathscr{F}_t)\right] \simeq \log(1-p_{k,j}) + d_k(Y)\log(1-p_{k,j}\omega).$$

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

$$-\Pr(y_{j,k}=1|\mathscr{F}_t) \simeq -p_{k,j} - d_k(Y)p_{k,j}\omega,$$

which implies

$$p_{k,j} \simeq \frac{\Pr(y_{jk}=1|\mathscr{F}_t)}{1+\omega d_k(Y)}. \tag{B.2.1}$$

The denominator evens out the 2nd stage's added attachment probability to nodes with higher degrees. I use (B.2.1) to readily compute the 1st-stage probabilities.

A comparison can be made with the alternative specification where one uses the right hand side of (2.4.2) 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 my 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.

B.3 Details on Posterior Computation

For the exposition of this subsection I will fix a time point t. Use R to denote the set of released movies: $\{k : r_k < t\}$, and Q the set of the yet to be released movies: $\{k : a_k \leq t, r_k \geq t\}$. Use $k \sim j$ to indicate that k and j are linked in the network. In my model the entire path up until t consists of $\mathscr{F}_t$ and ξ_Q. It is not difficult to see that the probability of the entire path up until t can be written as

$$\Pr\left(\mathscr{F}_t, \xi_Q\right) = \Psi(\mathscr{F}_t) \cdot \prod_{k \in Q \cup R} \Pr\left(\xi_k | y_k, \mathscr{S}_{a_k}\right).$$

The product term includes the arrival probabilities of the latent qualities. $\Psi(\mathscr{F}_t)$ 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 Ψ is a function of $\mathscr{F}_t$ only.

Given the specification in (2.4.3), 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 $\mathscr{F}_t$, we have

$$g\left(\xi_{Q \cup R}; \mathscr{F}_t\right) \equiv \prod_{k: a_k \leq t} \Pr\left(\xi_k | y_k, \mathscr{S}_{a_k}\right).$$

Then by the definition of conditional density, we have

$$\begin{aligned}
\Pr\left(\xi_Q | \mathscr{F}_t\right) &= \Pr\left(\mathscr{F}_t, \xi_Q\right) \cdot \left[\int \Pr\left(\mathscr{F}_t, \xi_Q\right) d\xi_Q\right]^{-1} \\
&= \Psi(\mathscr{F}_t) g\left(\xi_{Q \cup R}; \mathscr{F}_t\right) \cdot \left[\Psi(\mathscr{F}_t) \int g\left(\xi_{Q \cup R}; \mathscr{F}_t\right) d\xi_Q\right]^{-1} \\
&= g\left(\xi_Q | \xi_R; \mathscr{F}_t\right).
\end{aligned}$$

Given the specification in (2.4.3), one representation of the unconditional density g is

$$\xi_k = \sum_{\ell: a_\ell \leq t} W_{k\ell} \xi_\ell + \upsilon_k.$$

where $\upsilon_k \sim \mathcal{N}(0, V_{kk})$. 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:

$$\begin{aligned} W_{k\ell} &= \frac{\lambda\varphi^{|r_k-r_\ell|}}{1+\lambda\sum_{k\sim\ell,a_k<a_\ell}\varphi^{|r_k-r_\ell|}}, \text{ if } \ell\sim k \text{ and } a_\ell<a_k, \\ V_{kk} &= \frac{\sigma^2}{1+\lambda\sum_{k\sim\ell,a_k<a_\ell}\varphi^{|r_k-r_\ell|}}. \end{aligned}$$

In matrix form we can write

$$\xi_Q = W_{QR}\xi_R + W_{QQ}\xi_Q + \upsilon_Q,$$

or

$$\xi_Q = (I-W_{QQ})^{-1}(W_{QR}\xi_R + \upsilon_Q).$$

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

$$\Pr\left(\xi_Q|\mathscr{F}_t\right) = \mathscr{N}\left((I-W_{QQ})^{-1}W_{QR}\xi_R,\ (I-W_{QQ})^{-1}V_{QQ}(I-W'_{QQ})^{-1}\right)$$

To calculate the posterior on any subset $O\subseteq Q$, one 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 (i) Q is replaced with the subset of Q whose nodes are not separated from O by R, and (ii) R is replaced with the subset of R whose nodes are directly linked to some node in this subset 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 reduces to (2.4.3), the arrival distribution of j.

Table B.1: Monte Carlo Experiments for Supply Estimation

Parameters		Bias (%)	S.D. (%)
Attachment	2nd-stage Probability (ω)	1.0	1.2
	Intercept (γ_0)	0.7	0.4
	Own Movies (γ_1)	0.7	3.5
	Time Difference (γ_2)	0.5	2.0
Obs. Characteristics	Budget Mean w/o Precursors (μ)	4.2	12.7
	Budget Variance-to-Mean (χ)	-2.1	3.2
Coeff. of Risk Aversion (α)		3.6	16.3
Shock Size (ρ)		0.7	4.8
Yearly Arrival Rate (η)		-0.3	5.9

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.

B.4 Monte Carlo

I use Monte Carlo experiment to assess the supply-side estimator. The exercise amounts to simulating the model under the parameter point estimates to generate a dataset, and then applying the supply-side estimator to the dataset to recover the parameter values. This exercise is repeated a number of times to evaluate the distribution of the estimator. The results are displayed in Table B.1. All the parameters are recovered with an absolute bias smaller than 5%. The standard deviations of the estimator are used as the parametric bootstrapping standard errors for the supply-side estimates (see Table 2.6).

Bibliography

Chapter 1

Anselin, Luc, Raymond Florax and Sergio J. Rey (Eds.), 2004. "Advances in Spatial Econometrics: Methodology, Tools and Applications." Springer-Verlag, Heidelberg.

Aguirregabiria, Victor. and Chun-Yu Ho, 2011. "A Dynamic Oligopoly Game of the US Airline Industry: Estimation and policy experiments." *Journal of Econometrics*, vol. 168(1), 156–173.

Adler, Nicole, 2001. "Competition in A Deregulated Air Transportation Market." *European Journal of Operational Research*, 129: 337-345.

Abdelghany, Ahmed and Khaled Abdelghany, 2010. Chapter 1 in *Modeling Applications in the Airline Industry*. Ashgate Publishing.

Armantier, Olivier and Oliver Richard, 2008. "Domestic airline alliances and consumer welfare." *RAND Journal of Economics*, Vol.39 (3).

Berry, Steven, 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics*, Vol.25 (2), p.242-262.

Berry, Steven, Michael Carnall and Pablo T. Spiller, 2006. "Airline Hubbing, Costs and Demand." In *Advances in Airline Economics: Competition Policy and Antitrust, Vol. 1*, ed. Darin Lee, 183–214. Amsterdam: Elsevier.

Berry, Steven, James Levinsohn and Ariel Pakes, 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841-90.

Berry, Steven, Panle Jia, 2010. "Tracing the Woes: An Empirical Analysis of the Airline Industry." *American Economic Journal: Microeconomics*, Vol. 2(3), 1-43.

Borenstein, Severin, 1989. "Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry." *RAND Journal of Economics*, 20, 344-365.

Borenstein, Severin, 1992. "The Evolution of U.S. Airline Competition." *Journal of Economic Perspective*, Vol.6(2), p.45-73.

Borenstein, Severin, 2011. "Why Can't US Airlines Make Money?" *American Economic Review*, 101(3): 233-37.

Bramoullé, Yann, Habiba Djebbari and Bernard Fortin, 2009. "Identification of peer effects through social networks." *Journal of Econometrics*, 150(1): 41–55.

Brueckner, Jan K., and Pablo T. Spiller, 1994. "Economies of Traffic Density in the Deregulated Airline Industry." *Journal of Law and Economic*s, 37, 379-415.

Caves, Douglas, Laurits Christensen and Michael Tretheway, 1984. "Economies of Density Versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ." *RAND Journal of Economics*, 15, 471-489.

Ciliberto, Federico and Jonathan W. Williams, 2010. "Limited Access to Airport Facilities and Market Power in the Airline Industry." *Journal of Law and Economics*, 53(3), p.467-495.

Ciliberto, Federico and Jonathan W. Williams, 2014. "Does Multimarket Contact Facilitate Tacit Collusion? Inference on Conduct Parameters in the Airline Industry." *RAND Journal of Economics*, Vol. 45 (4).

Dobson, Gregory and Phillip J. Lederer, 1993. "Airline Scheduling and Routing in a Hub-and-Spoke System." *Transportation Science*, Vol.27(3).

Einav, Liran and Jonathan Levin, 2010. "Empirical Industrial Organization: A Progress Report." *Journal of Economic Perspectives*, 24(2):145–162.

Gastner, Michael T. and Mark E. J. Newman, 2006. "Optimal design of spatial distribution networks." *Physical Review E*, 74, 016117.

Gandhi, Amit, Zhentong Lu and Xiaoxia Shi, 2015. "Estimating Discrete Choice Models with Market Level Zeroes: An Application to Scanner Data." Working Paper, University of Wisconsin-Madison.

Gillen, David W., William G. Morrison, and Christopher Stewart, 2003. "Air Travel Demand Elasticities: Concepts, Issues and Measurement." Department of Finance Cananda Report.

Givoni, Moshe and Piet Rietveld, 2009. "Airline's Choice of Aircraft Size - Explanations and Implications." *Transportation Research Part A*, 43, 500-510.

Hansen, Mark and Yi Liu, 2015. "Airline competition and market frequency: A comparison of the s-curve and schedule delay models." *Transportation Research Part B*, Vol. 78, p. 301–317.

Hendricks, Ken, Michele Piccione, and Guofu Tan, 1999. "Equilibria in Networks." *Econometrica*, 67(6), 1407-1434.

Mazur, Joe, 2015. "Can Stricter Bankruptcy Laws Discipline Capital Investment? Evidence from the U.S. Airline Industry." Working Paper, Duke University.

Manski, Charles F., 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60(3): 531-542.

McFadden, Daniel, 1981. "Econometric Models of Probabilistic Choice." Structural Analysis of Discrete Data with Econometric Applications, 198– 272, MIT Press.

Nevo, Aviv, 2000. "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry." *RAND Journal of Economics*, 31, 395–421.

Norman, Victor D. and Siri Strandenes, 1994. "Deregulation of Scandinavian Airlines: A Case Study of the Oslo-Stockholm Route." Chapter 4 in Empirical Studies of Strategic Trade Policy, Paul Krugman and Alasdair Smith (Eds.), University of Chicago Press.

Peters, Craig, 2006. "Evaluating the Performance of Merger Simulation: Evidence from the U.S. Airline Industry." *Journal of Law and Economics*, vol. XLIX., pp 627-649.

Richard, Oliver, 2003. "Flight Frequency and Mergers in Airline Markets." *International Journal of Industrial Organization*, Vol. 21, 907-922.

Roberts, James W. and Andrew Sweeting, 2013, "Airline Mergers and the Potential Entry Defense." Working paper, Duke University.

Stock, James, Jonathan Wright and Motohiro Yogo, 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Buesiness & Economic Statistics*, Vol. 20(4).

Wei, Wenbin and Mark Hansen, 2005. "Impact of Aircraft Size and Seat Availability on Airlines' Demand and Market Share in Duopoly Markets." *Transportation Research Part E*, 41: 315–327.

Chapter 2

Aguirregabiria, Victor and Chun-Yu Ho, 2011. "A Dynamic Oligopoly Game of the US Airline Industry: Estimation and Policy Experiments." *Journal of Econometrics*, Vol. 168(1).

Alberts, Bruce, 2012, "The End of Small Sciences?" *Science*, Vol. 337, p. 1583.

Argote, Linda and Dennis Epple, 1990, "Learning Curves in Manufacturing." *Science*, Vol. 247, p. 920-924.

Andersen, Torben and Bent Sørensen, 1996. "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study." *Journal of Business & Economic Statistics*, Vol. 14(3), pp. 328-352.

Ainslie, Andrew, Xavier Drèze and Fred Zufryden, 2005. "Modeling Movie Life Cycles and Market Share." *Marketing Science*, Vol. 24 (3), 508-17.

Barabási, Albert-László and Réka Albert, 1999. "Emergence of scaling in random networks." *Science*, Vol. 286, pp.509–512.

Benkard, Lanier, 2000. "Learning and Forgetting: The Dynamics of Aircraft Production." *American Economic Review*, 90(4), 1034-1054.

Bradlow, Eric, et al. 2005, "Spatial Models in Marketing." *Marketing Letters*, 16 (3-4), 267 - 278

Chintagunta, Pradeep, 1994, "Heterogeneous Logit Model Implications for Brand Positioning." *Journal of Marketing Research*, Vol. 31, pp. 304-311.

Ching, Andrew T., Tülin Erdem and Michael P. Keane, 2013. "Learning Models: An Assessment of Progress, Challenges, and New Developments." Marketing Science, Vol. 32(6).

De Vany, Arthur and David Walls, 1996. "Bose-Einstein Dynamics and Adaptive Contracting in the Motion Picture Industry." *The Economic Journal*, Vol. 106, pp. 1493-1514

De Vany, Arthur and David Walls, 1999, "Uncertainty in the movie industry: Does star power reduce the terrorof the box office?" *Journal of Cultural Economics*, Vol. 23: 285–318

Dellarocas, Chris, Zsolt Katona, and William Rand, 2010. "Media, aggregators, and the link economy: Strategic hyperlink formation in content networks." Working Paper, NET Institute, New York.

Desrosiers, Christian and George Karypis, 2011, "A Comprehensive Survey of Neighborhood-based Recommendation Methods," *Recommender Systems Handbook*, pp. 107-144.

Einav, Liran, 2007. "Seasonality in the U.S. Motion Picture Industry." *RAND Journal of Economics*, 38(1).

Einav, Liran, 2010. "Not All Rivals Look Alike: Estimating an Equilibrium Model of The Release Date Timing Game." *Economic Inquiry*, 48(2).

Eizenberg, Alon, 2014. "Upstream Innovation and Product Variety in the U.S. Home PC Market." *Review of Economic Studies*, Vol.81, 1003–1045

Elberse, Anita, and Jehoshua Eliashberg, 2003. "Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures." *Marketing Science*, Vol. 22(3).

Eliashberg, Jehoshua, Anita Elberse and Mark Leenders, 2006, "The Motion Picture Industry: Critical Issues in Practice, Current Research, and New Research Directions." *Marketing Science*, Vol. 25(6).

Eliashberg, Jehoshua, Sam Hui, and John Zhang, 2007. "From Story Line to Box Office: A New Approach for Green-Lighting Movie Scripts." *Management Science*, 53 (6), 881 - 893.

Erdös, Paul and Alfréd Rényi, 1959. "On Random Graphs. I" *Publicationes Mathematicae*, 6:290-297.

Elrod, Terry and Michael P. Keane, 1995, "A Factor-Anaiytic Probit Model for Representing the Market Structure in Panel Data." *Journal of Marketing Research*, Vol. 32, pp. 1-16.

Goettler, Ronald and Phillip Leslie, 2005. "Cofinancing to Manage Risk in the Motion Picture Industry." *Journal of Economics & Management Strategy*, 14(2).

Goettler, Ronald and Ron Shachar, 2001. "Spatial Competition in the Network Television Industry." *RAND Journal of Economics*, Vol. 32(4), 624-656.

Goldenberg, Jacob, Gal Oestreicher-Singer and Shachar Reuchman, 2012. "The Quest for Content: How User-Generated Links Can Facilitate Online Exploration." *Journal of Marketing Research*, 49(4), p. 452-468.

Han, Chirok and Peter C.B. Phillips, 2006. "GMM with Many Moment Conditions." *Econometrica*, Vol. 74(1), 147–192

Hennig-Thurau, Thorsten, Mark B. Houston and Shrihari Sridhar, 2006, "Can good marketing carry a bad product? Evidence from the motion picture industry." *Marketing Letters*, 17: 205-219.

Hitsch, Günter J., 2006. "An Empirical Model of Optimal Dynamic Product Launch and Exit Under Demand Uncertainty." Marketing Science, 25 (1).

Holme, Petter and Beon Jun Kim, 2002. "Growing scale-free networks with tunable clustering." *Physical Review: E*, 65(2).

Irwin, Douglas A. and Peter J. Klenow, 1994. "Learning-by-Doing Spillovers in the Semiconductor Industry." *Journal of Political Economy*, Vol. 102(6), p.1200-1227.

Jackson, Matthew O., 2010. Chapter 5 in *Social and Economic Networks*. Princeton University Press.

Jackson, Matthew O., and Brian Rogers, 2005. "Meeting Strangers and Friends of Friends: How Random Are Social Networks?" *American Economic Reivew*, 97(3): 890-915

Lambert, Richard A., 1986, "Executive Effort and Selection of Risky Projects." *RAND Journal of Economics*, 17(1), 77–88.

LeSage, James, 2008, "An Introduction to Spatial Econometrics." *Revue d'économie industrielle*, Vol. 123, p.19-44.

Linden, Greg, Brent Smith and Jeremy York, 2003. "Amazon.com Recommendations: Item-to-Item Collaborative Filtering." *IEEE Internet Computing*, Vol 7(1), 76-80.

Luo, Hong, 2014. "When to Sell Your Idea: Theory and Evidence from the Movie Industry." *Management Science*, 60 (12), 3067-3086.

Mayzlin, Dina and Hema Yoganarasimhan, 2012. "Link to success: How Blogs Build an Audience by Promoting Rivals." *Management Science*, 58(9): 1651-1668.

Mednick, Sarnoff, 1962, "The Associative Basis of The Creative Process." *Psychological Review*, Vol. 69 (3).

Newman, Mark, 2003, "The Structure and Function of Complex Networks." *SIAM Review*, Vol. 45(2).

Oestreicher-Singer, Gal and Arun Sundararajan, 2012. "The Visible Hand? Demand Effects of Recommendation Networks in Electronic Markets." *Marketing Science*, 58: 1963–1981.

Orbach, Barak and Liran Einav, 2007, "Uniform Prices for Differentiated Goods: The Case of the Movie-Theater Industry." *International Review of Law and Economics*, 27(2).

Pesendorfer, Wolfgang, 1995, "Design Innovation and Fashion Cycles." *American Economic Review*, Vol. 85(4).

Prag, Jay, and James Casavant, 1994, "An Empirical Study of the Determinants of Revenues and Marketing Expenditures in the Motion Picture Industry." *Journal of Cultural Economics*, Vol. 18(3).

Raustiala, Kal and Christopher Jon Sprigman, 2006, "The Piracy Paradox: Innovation and Intellectual Property in Fashion Design." *Virginia Law Review*, Vol. 92, p. 1687.

Ravid, S. Abraham, 1999. "Information, Blockbusters, and Stars: A Study of the Film Industry." *Journal of Business*, 72(4).

Rue, Havard and Leonhard Held, 2005. *Gaussian Markov Random Fields: Theory and Applications*. CRC Press.

Shen, Qiaowei and Ping Xiao, 2014a. "McDonald's and KFC in China: Competitors or Companion." *Marketing Science*, 33(2).

Shen, Qiaowei and Hongju Liu, 2014b, "Demand Uncertainty, Dynamic Learning and Exit in Competitive Markets." Working Paper, University of Pennsylvania.

Squire, Jason, 2005. *The Movie Business Book*, 3rd Edition. Simon and Schuster.

Schmidhuber, Jürgen, 2011, "Citation Bubble About to Burst?" *Nature*, Vol.469 (34).

Toivanen, Otto. and Michael Waterson, 2005. "Market Structure and Entry: Where's the Beef?" *RAND Journal of Economics*, 36, 680-699.

Thornton, Rebecca A. and Peter Thompson, 2001. "Learning from Experience and Learning from Others: An Exploration of Learning and Spillovers in Wartime Shipbuilding." *American Economic Reivew*, Vol. 91(5).

Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Ben Jones, 2013. "Atypical Combinations and Scientific Impact." *Science*, Vol.342: 468-472.

Wallace, Timothy, Alan Seigerman, and Morris Holbrook, 1993. "The Role of Actors and Actresses in the Success of Films: How Much Is a Movie Star Worth?" *Journal of Cultural Economics*, 17(1), 1-27.

Watts, Duncan J. and Steven Strogatz H. 1998. "Collective Dynamics of 'Small-World' Networks." *Nature*, 393 (6684).

Weitzman, Martin L., 1998. "Recombinant Growth." *Quarterly Journal of Economics*, Vol.113, 331-360.

Wei, Yanhao, 2014. "Network Effects of Air-Travel Demand." Working Paper No. 14-027, Penn Institute for Economic Research.

Yang, Nathan, 2014. "March of the Chains: Herding in Restaurant Locations." Working Paper, Yale University.