

ABSTRACT

The Long Tail: An Evaluation of the Causal Assumptions in Market Determination

Troy Lee Hales II

Director: Dr. Gary Carini

The “Long Tail,” is a market distribution wherein obscure products sell in large aggregate amounts, allowing them to compete comparatively well with the few popular products in the same market that sell well individually. This type of market came to the forefront of business research in a 2006 trade book by *Wired* editor-in-chief, Chris Anderson, titled *The Long Tail*. Focusing on the example of the book publishing industry in the United States, the paper explores each of Anderson’s three assumptions he claims are the causes necessary to the emergence of the Long Tail in a market—the “democratization of production,” the “democratization of distribution,” and supply filters. Utilizing the Counterfactual framework established by Morgan and Winship, which draws from Judea Pearl’s Directed Acyclic Graphical model, this paper displays Anderson’s assumptions in conjunction with additional considerations absent in Anderson’s framework.

APPROVED BY DIRECTOR OF HONORS THESIS:

Dr. Gary Carini, Professor and Associate Dean for Graduate
Business Programs

APPROVED BY THE HONORS PROGRAM:

Dr. Andrew Wisely, Director

DATE: _____

*THE LONG TAIL: AN EVALUATION OF THE CAUSAL ASSUMPTIONS IN MARKET
DETERMINATION*

A Thesis Submitted to the Faculty of
Baylor University
In Partial Fulfillment of the Requirements for the
Honors Program

By
Troy Lee Hales II
Waco, TX
April 2013

TABLE OF CONTENTS

| | |
|----------------------------|-----|
| LIST OF FIGURES AND TABLES | III |
| CHAPTER ONE | 1 |
| CHAPTER TWO | 12 |
| CHAPTER THREE | 22 |
| CHAPTER FOUR | 29 |
| CHAPTER FIVE | 39 |
| BIBLIOGRAPHY | 42 |

LIST OF FIGURES AND TABLES

| | |
|-----------------|----|
| Figure 1.0..... | 4 |
| Figure 1.1..... | 9 |
| Figure 4.0..... | 29 |
| Figure 4.1..... | 31 |
| Table 4.0..... | 32 |
| Figure 4.2..... | 32 |
| Figure 4.3..... | 34 |
| Figure 4.4..... | 35 |
| Figure 4.5..... | 36 |
| Figure 4.6..... | 37 |

CHAPTER ONE

Introduction

The Internet has, over the past two decades, had a significant influence over business practices and the economy. This is evident in many facets of commerce: pre-recorded music and film; the news media; advertising. In general, the Internet simplifies many of the critical components of profit making. Whether a company is looking to reduce the costs of distribution, seeking a cheaper alternative to current production methods, or simply trying to establish a connection with customers, the Internet is a popular tool with which to accomplish the end of more fluid, diverse business.

Research surrounding the effects of the Internet on markets has, in turn, surged in the past two decades—when it began to assume a more significant role in business processes¹. Of particular interest is a line of literature beginning with an article published in 2004 titled “The Long Tail,” which debuted in *Wired* magazine². In it, Chris Anderson, *Wired*’s chief editor, describes a type of market distribution wherein “niche products” are a force with which to be reckoned. In a market with a “long tail,” he says, we find “an entirely new economic model.”

¹ Lovelock, *Forecast Alert: IT Spending, Worldwide, 4Q12 Update* | 2291618.

² Anderson, “Wired 12.10: The Long Tail.”

He articulates three main “rules” that he claims lead to the emergence of a long tail. In his 2006 book, *The Long Tail: Why the Future of Business is Selling Less of More*, he develops these rules further, into conditions for long tail emergence. In markets where producers are “democratizing production,” “democratizing distribution,” and bridging the gap between supply and demand, a long tail will certainly surface³. In evaluating the literature since 2003 as well as developing a possible research agenda to explore these assumptions in the United States book publishing industry, this paper hopes to show that the causal inferences Anderson makes are flawed for being causal.

Scope of the Paper

This paper evaluates Anderson’s three assumptions within the directed acyclic graphical (DAG) model developed in Morgan and Winship’s work, *Counterfactuals and Causal Inference*, in order to show that the assumptions are weak when considered alone, as they are in Anderson’s construct.

The remainder of this chapter defines the “long tail effect” more fully and explores the differences between traditional and long tail markets; it explains the equally relevant but opposite “superstar effect” that the existing literature explores; and it discusses the importance of a model which could predict the emergence of long tails in online markets.

The second chapter summarizes the literature pertaining to the long tail and superstar effects and elucidates the application of the two to various types of markets. Very little, if any, scholarship exists on the validity of Anderson’s claims as

³ Anderson, *The Long Tail*.

causal.

The third chapter begins by identifying the characteristics of the book publishing industry and proceeds to investigate the history of the United States publishing industry, in particular. Next, the paper highlights facets of the book publishing industry that could reasonably contribute to the “long tail effect” in addition to the three assumptions Anderson discusses.

The fourth and fifth chapters develop a DAG model that could be used as a more definite guide to researching the emergence of a long tail in the U.S. book publishing industry and test it in a hypothetical business. This portion intends to show that more nuanced research than what Anderson conducts is necessary to make valid inference in long tail prediction.

What is the “Long Tail?”

Description

The name of the distribution, itself, is merely a description of the graphical representation of a certain type of market. Alone, this title does not describe the concept well. To restate the abstract a bit is necessary to establish both conceptually and pictorially what a long tail market is.

Generally, a representation of market distribution illustrates sales as the dependent variable (in quantity or monetary units on the y-axis) and number of products as the independent variable (x-axis). As the amount of products in a market increases, we see the “tail” of the graph get longer and, conversely, as the sales of the individual products increase, we see the “head” become steeper. Most

representations organize the products in sales of descending order; specifically, the individual product with the most sales is closer to the y-axis than any other and, oppositely, the individual product with the least sales is farthest from the y-axis than any other.

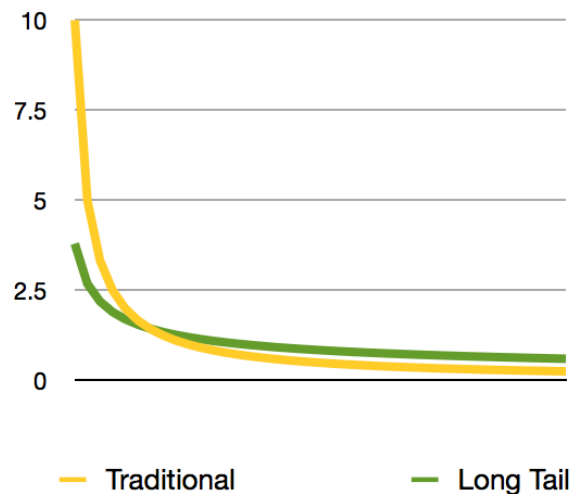


Figure 1.0

This same construct works with number of firms as an independent variable and some factor like the popularity of each respective firm as a dependent variable. If we take the industry for raw, packaged meat as an example, we would plot popular firms, such as WalMart, nearest the y-axis and less popular firms, such as a neighborhood butcher, further from that axis in the “tail.” In the book publishing industry, which we expand on in great detail in Chapter 3, one may identify each publishing entity as a point on the ranked x-axis. Then the more famous publishers like Random House and Cengage would appear very near the y-axis and less popular niche writers publishing their own work appear far into the tail.

The most popular products aggregate nearest the y-axis in what is called the

“head” of the distribution. As individual popularity declines, products begin to occupy the “tail” distribution, though there exists no clear point at which products are considered part of the “tail,” according to Anderson. In a traditional market, where mainly popular products are sold, there is a large concentration of products in the “head” of the graph. On the other hand, long tail markets have a larger concentration of products in the “tail,” relative to the traditional market for comparable products.

The illustration above (Figure 1.0) will perhaps clarify the concept. Firms are ranked by level of sales, and we assume a continuous amount of firms in the market so that the graph is smooth.

Both markets boast the same amount of sales, but it is obvious that concentration is distributed more widely in the long tail. The traditional market, in gold, has a larger share of products concentrating in the head of the distribution. This means that fewer products sell in great amounts—Anderson refers to these popular products as “blockbusters.” The long-tailed market, in green, tells a different story. The tail in this market does not quite approach zero as quickly as the traditional market’s does. This means that the less popular niche products not only sell in greater amounts individually than in regular markets, there are *more* of such individual products available. This implication is intriguing as it shows that in long-tailed markets, there are greater numbers of autonomous products supplied and, more importantly, that consumers actually purchase them. Anderson argues that this is because there is latent demand that is untapped in the traditional market that is captured online.

Emergence Assumptions

The most basic of assumptions Anderson makes in his theory of long tail emergence is that of the presence of new technology. The price of servers decreased dramatically over the past few decades, which has had the implication of decreasing the cost of creating and hosting content online for nearly everyone with access to a computer. Simultaneous advances in cheaper, more efficient, and expansive computing power and storage for end users (as well as firms and expert users) has led to increasing interaction with all of the new content. Not only do most Americans own a computer, cell phone, or tablet (or have easy access to any of the three), those devices pack more power in significantly less space than before and are only getting better.

With such power available to them, end users become more involved in the production of content. According to Anderson, this is the first of three assumptions that—when taken together—illicit a “long tail effect.” When users create content at home and for no prescribed benefit, they contribute to the “democratization of production” within a market. Most often, they create niche products, appealing only to small audiences. Occasionally, a user ends up creating a product that is popular enough to become a “blockbuster.” It is important to note, however, that this is the exception. As Anderson says, “there’s [a] lot of crap” created by amateur users, but the mere fact that they *are* creating is crucial. He notes, “more than half of Amazon’s book sales come from *outside* its top 130,000 titles,” and that “[t]he average Barnes & Noble carries 130,000 titles.” His argument is that amateur content makes up much of the products (and sales) outside those top titles that traditional markets

will carry at the exclusion of niche products. Conflicting literature contends that, though the user-created content does exist in online markets, it does not make up a valuable portion of sales in most cases⁴.

If this is true—that user-created products do not in general make a useful contribution to the market—it is important to inquire as to why they are even there to begin with. It is at this point that we reach the first large gap in Anderson’s model—the assumption of low barriers to entry into the market. With low barriers to entry, users have little disincentive to placing their “crap” products in the market. Without, they are forced to contend with such factors as cost constraints and higher risk. Anderson instead focuses on the consequent of low barriers, what he calls “democratization of distribution.” Users can place content straight onto sites like Amazon or Youtube for others to observe, review, or ignore. The Internet makes this possible but does not guarantee the degree of disintermediation that Anderson proclaims. Certain markets, like that for legal services, remain heavy in the head of the distribution because high barriers to entry are imposed for legitimate, expertly rendered products.

An improved connection of supply to latent demand is the last of Anderson’s assumptions. Online markets must boast some mechanism by which they allow consumers to filter products and better increase the probability of finding the product(s) they seek. User review capability is also quite essential to the growth of a long tail. This provides an additional filter for the content that users may actively engage. Anderson does not provide a threshold for this assumption—so any

⁴ Elberse and Oberholzer-Gee, “Superstars and Underdogs.”

functional search feature would qualify as meeting this third condition. This weak criterion is the point at which most divergence occurs in the literature. Do all search features contribute to long tail emergence, or do some have an opposite effect? The literature argues that varying degrees and types of search technology have a significant impact upon whether a distribution is long-tailed.

What is the “Superstar Effect?”

Description

William McPhee’s standing theories of “natural monopoly” and “double jeopardy,” introduced in 1963, are in direct contradiction to Anderson’s thoughts on latent demand. In essence, McPhee’s work proposes that people tend to stick with the familiar and, even when they do venture into the long tail, they would not be happy enough with the products there to stay⁵. The implication of McPhee’s thoughts on people’s behavior is, when applied to economic demand, in direct contradiction with Anderson’s idea that there is a large amount of demand not met by traditional markets. Where Anderson claims there are plenty of consumers looking for unique niche mystery serials on Amazon, McPhee would say that there are many more in search of hit literature. Though Anderson shows that this is not always the case, there is merit in exploring the implication of McPhee’s theory in today’s online markets.

The superstar effect, many contend, arises following many of the same criteria of the emergence of a long tail. The result is a greater magnitude of sales of

⁵ McPhee, *Formal Theories of Mass Behavior*.

hit products than usual. An online market may contain many products more than a traditional market, but it also boasts many more consumers, in general. When those consumers choose to purchase what they are comfortable with, which McPhee's theory would posit, the hits that they purchased in distinct markets are now aggregated in a single market that has centralized feedback mechanisms and search capabilities. Consumers can often not only see what other consumers think of a product and speak into that opinion themselves, they can also see *who* is buying that product. When the right person or group of people begins to buy products, others are drawn to purchase those products, as well. We can see why this would be effective by noting the use of celebrities and other famous figures in television advertising.

When a positive feedback loop initiates in an online market, the superstar effect (whose result is illustrated below in Figure 1.1) can begin to define market distribution.

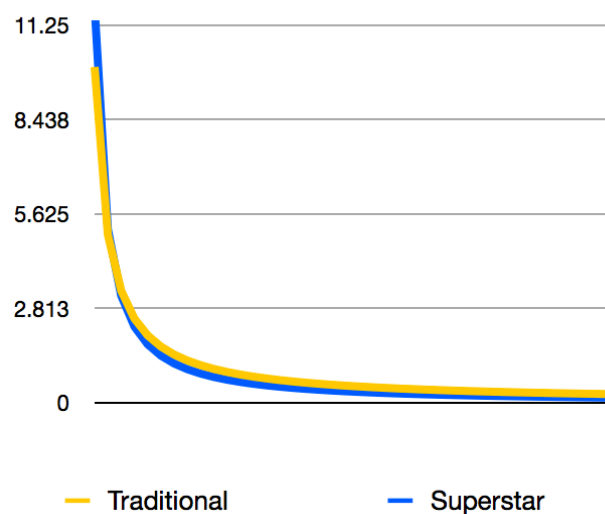


Figure 1.1

As is observed in Figure 1.0, both markets have the same amount of sales, in total. However, in this illustration (though it may be more difficult to see) it is the case that the “superstar” market has more sales concentrated in the head of the distribution. Less niche products sell than in the traditional market and more hit products flow from producers to the eager consumers.

Emergence Assumptions

The first two of Anderson’s emergence assumptions, disintermediation of both production and distribution, are described sufficiently above and are similarly applied in the literature to the conditioning of the superstar effect. Therefore, we begin by examining the differences in application of the third assumption.

Not only do many of the critics of Anderson’s theory introduce McPhee’s ideology to diminish the argument for a pre-existing latent demand, they also commonly note that only certain types of search mechanisms tend to produce long tails. Rather—many claim—a superstar effect is more common due to firm influence on search results. Companies have control over the mechanisms by which consumers discover products in most cases; incentives exist to tailor the results of a search toward the exchange of products from which the company would expect to most benefit. Anderson’s third assumption, then, crumbles without proper specification.

What is the Value of such a Model?

Markets are moving online at a rapid rate in most industries⁶. Undeniably, business practices, consumer preferences, and educational programs will continue to be affected. Whether a long tail or superstar market should be expected to emerge is an increasingly relevant question for research.

Businesses will have an interest in the answer to this when developing Internet marketing strategies. Given the current state of an online market, should a firm that produces hit products enter, or would a niche producer fare better? If a company can predict with confidence that profit would be negative in certain online markets, it may choose to introduce its product elsewhere, or pursue a different product, entirely. In either case, the ability to more accurately predict long tail emergence enters directly into a firm's profit-maximization concerns.

Long tail emergence is also of great relevance to regulatory interests. This is not obvious in the case of the online market for used books, for example, where substandard quality is more of a nuisance than a significant problem. The same cannot be said for an online market for legal services, however. Improperly rendered or, in many cases, "do it yourself" legal services can be detrimental to personal welfare. With the ability to predict the emergence of long tails, regulatory bodies can more proactively prevent misleading or substandard products from ever reaching consumers.

⁶ Crowdhry, *Internet As A Building Block*.

CHAPTER TWO

Literature Review

The earliest literature on the emergence of the long tail is relatively young. Though some seminal, relevant consumer behavior articles this paper shall discuss date back to the late fifties and early eighties, published academic writing on this subject began as late as 2003. Since then, it has developed through four distinct stages. The foundational work from 2003 to around 2006 introduced the concept of the long tail. Definitional work followed, building models within which to evaluate the phenomenon. Simultaneously, a debate amongst researchers began concerning the future of business being better categorized as one belonging to superstars or long tails. Last, particularly in the past few years, researchers worked to integrate these opposing effects into singular, more specific, predictive models.

Foundational Work

The Long Tail theory is one born of conflict. Anderson coined the term in 2004 in an article published by *Wired* magazine. In it, he provides some examples of information goods exhibiting “long tail behavior” and gives three “rules” for managers to follow in order to take advantage of online markets⁷. He says that, at least in the entertainment industry, future business success will be contingent upon

⁷ Anderson, “Wired 12.10: The Long Tail.”

“mak[ing] everything available, [...] cut[ting] the price, [and] help[ing consumers] find [products].” If logically extended a few steps, Anderson’s claims can simplify to say that cheap, highly available and diverse, filtered products will naturally organize themselves into a long tail market. Similar conditions, discussed in an earlier work by Clay Shirky (2003), may be extended in the opposite direction.

Freedom in the weblog world, plenty of blogs updated often, and filtration systems are Shirky’s conditions for the emergence of a powerlaw, or superstar, market⁸. Shirky contends that powerlaw distributions will naturally arise wherever these conditions exist, given that the filtration system for consumers is one based upon recommendation. Anderson does not make this pivotal distinction between recommendations and other search tools or passive instruments used by consumers or producers to filter supply. This is perhaps the largest point of contention in the literature in and after 2006. Shirky also begs an interesting question that evades Anderson in his writings: where does the long tail start? The head of the distribution is composed of blockbusters and the tail of niches—neither Anderson nor Shirky establish where these two categories become distinct. This discrepancy is explored definitively almost seven years later, in the period of definitional long tail literature.

Another key distinction Anderson does not draw out is that of product variety. An earlier piece by Brynjolfsson, Hu, and Smith (2003) supports Anderson’s technology assumption but advances the idea that product variety is key in the emergence of a long tail. Though this is an implicit assumption Anderson makes in

⁸ “Shirky: Power Laws, Weblogs, and Inequality.”

his book two years later, it is worth isolating as its own piece of the model.

Brynjolfsson, Hu, and Smith argue that consumer surplus improves through this increased product variety, which may only come about with less than prohibitive transactions costs as well as with recommendation and search tools⁹.

Anderson follows this previous literature by developing the three main assumptions discussed in this paper's introductory chapter: disintermediation of both production and distribution and accompanying supply filtration mechanisms within the market. He also introduces the idea of latent demand, claiming that "the economics of abundance" allow consumers to unveil "the true shape of demand" for themselves. His observations point out that consumers may have more nuanced preferences than they previously thought when confronted with more limited choices.

The final portion of foundational literature this paper identifies is another piece from Brynjolfsson, Hu, and Smith that extends Anderson's discussion of the "drivers" of an emerging long tail. They contend that lower storage and production costs, easier access to promotional and distribution channels, and active or passive search mechanisms promote long tail emergence¹⁰. Importantly, Brynjolfsson, *et al*, begin developing a definitional system that they later expand upon wherein the conditions of long tail emergence may be classified as either demand or supply side factors.

⁹ Brynjolfsson, Yu Hu, and Smith, "Consumer Surplus in the Digital Economy."

¹⁰ Brynjolfsson, Hu, and Smith, "From Niches to Riches."

Definitional Work

The long tail, well established as a market structure common in online markets following Anderson's and Brynjolfsson's, *et al*, observations, still evades clear definition. Up to this point, the literature discussed potential, non-empirical foundations for recognizing a long tail before or during its emergence. These foundations do not address the discrepancies about filtration mechanisms, Shirky's question of where the long tail begins, or how researchers might actually progress in studying the long tail with some cohesive model.

Filtration systems and, in particular, recommendation networks are the nexus of long tail and superstar effect emergence. Fleder and Hosanagar discuss the three implications they observe to arise from the use of recommendation systems. Their research suggests that these networks tend to (i) either make the rich get richer, the poor get poorer, or both; (ii) increase individual-level diversity but usually decrease aggregate diversity; and (iii) have a bias toward certain products created by the design of the system¹¹.

Their paper is but one of many implying that the filtration system assumption introduced by Anderson is faulty. Because this purported necessary condition is applicable to both the Superstar effect and the emergence of a Long Tail, it is a mistake to define the model of Long Tail emergence on this assumption without further specifying its parameters. This paper narrows Anderson's broader assumption for a more limited condition introduced in the literature to come.

¹¹ Fleder and Hosanagar, *Blockbuster Culture's Next Rise or Fall*.

In fact, work by Clemons and Gao (2008) posits that online markets, in general, mitigate a blight of traditional markets called “uncertainty discount.” Consumers may not only search for the products they desire, they may also read and process what others say about those products, as well. Clemons and Gao conceptually argue that recommendation systems made to inform consumers tend to lead them to find products better matching the consumers’ preferences. This propagates the long tail. On the other hand, the authors imply, recommendation and search systems serving interests other than “consumer informedness” pull markets away from becoming a long tail. Therefore, the research suggests that long tail emergence requires not just search and recommendation mechanisms, but those that contribute to consumers’ understanding of available products and services¹².

With Anderson’s third assumption better specified, the literature begins to focus on defining the point in a market at which products change from “blockbuster” to “niche” classification. In response to the myriad articles published concerning the long tail, all lacking a cohesive model, Brynjolfsson, Hu, and Smith discuss the importance of such a model, what it might look like, and what the literature has produced in the meantime¹³.

The authors begin by reiterating that the long tail is here to stay, suggesting it is a relevant study for both academics and businesspeople; this is echoed in more recent research, as well¹⁴. They then establish a framework for classifying the drivers responsible for the emergence of long tails (or Superstars, where the model

¹² Clemons and Gao, “Market Transformation in a Networked Global Economy.”

¹³ Brynjolfsson, Yu Hu, and Smith, “Long Tails Vs. Superstars.”

¹⁴ Peltier and Moreau, “Internet and the ‘Long Tail Versus Superstar Effect’ Debate.”

predicts their emergence). Drivers fall into one of four categories: demand-side technological or non-technological and supply-side technological or non-technological. Papers succeeding Brynjolfsson, *et al's* (2006) work tend to utilize this terminology and, in general, focus on demand-side technological drivers like recommendation systems or on demand-side non-technological drivers like consumer desire for social interaction.

Brynjolfsson, *et al*, (2010) also evaluate existing literature's many models for the long tail, pointing out the strengths and weaknesses of each. They claim that, "the long tail can be defined and measured in at least three different ways," and that researchers should probably choose from one of these and discuss why they chose a particular model. I focus on the two most relevant, here, as the third does not arise in the literature before or after this 2010 paper very often.

The "Absolute Long Tail" sets a cutoff point at some absolute quantity. Though he does not use this terminology, Anderson's example of the long tail—whichever products (however many there may be) do not make it into a Barnes & Noble are part of the long tail—is the absolute long tail. Whether that quantity is set at 1,000 products or 100,000 is entirely at the discretion of the researcher and can often lead to inconsistencies when comparing distributions in different markets. A benefit, however, is the intuitive simplicity—if, in Anderson's model, a book cannot be found at Walmart or Barnes & Noble, it belongs to the long tail. Researchers establish the "Relative Long Tail" in similarly arbitrary fashion. By choosing a cutoff percentage such as 20%, as with the famous Pareto Principle, researchers may compare distributions more freely across markets but still fall prey to inconsistency

if another researcher chooses, for example, 30% as the cutoff point. Importantly, researchers begin to note and defend the use of particular cutoffs following this specification. Though this did not establish a completely cohesive model for long tail emergence, it helps to compare and discuss further literature on the subject.

The Division

In 2008 a debate begins to emerge amongst researchers about whether the future of business really is, as Anderson puts it, selling less of more. The vigorous arguments for the superstar effect usually begin by citing the non-technological supply factors of consumer attraction to popularity and talent. Superstar proponents contend research shows that consumer preferences align often because people are naturally drawn to better quality products or more talented service providers¹⁵. In addition, consumers like to enjoy and think about products or services with one another; purchasing is a social behavior¹⁶. Therefore, consumers will use online markets to get more of what they already like...blockbusters.

Other sources continue with Anderson's original argument that niche products will tend to divert consumers from the hits they have been purchasing in traditional markets. Entertainment industries are experiencing wider demand distributions, some research contends, so why would other online markets not exhibit similar behavior¹⁷?

¹⁵ Rosen, "The Economics of Superstars."

¹⁶ McPhee, *Formal Theories of Mass Behavior*.

¹⁷ *Black Book - IP Video*.

As this and other literature begins to emerge both in favor and against Anderson's observations, Elberse (2008) publishes the article "Should You Really Invest in the Long Tail?" She cites some examples of the superstar effect arising under Anderson's conditions for long tail emergence and concludes that the discerning manager should be wary of Anderson's suggestions¹⁸. This sparks an actual debate between the two on the *Harvard Business Review* blogs. Anderson contends that his original predictions are correct and that Elberse simply defines the long tail differently than he does. Her response concludes that his definition is arbitrary¹⁹. This disagreement seems to incite Elberse to action, however, because she soon publishes an article with Oberholzer-Gee surveying the digital video sales market for a long tail.

The results show that fewer products belong in the blockbuster category within a few years of the market moving online, but that the small amount that remain continue to sell very well. They also show a wider but flatter long tail. More niche products exist and some tend to sell better, though most do not. So, Elberse and Oberholzer-Gee conclude that there are both superstar and long tail effects present in the market but that it would be inadvisable to fully commit to one or the other currently. In fact, they claim that, though niche products are more profitable for businesses than they were previously, businesses should avoid them until online markets are better understood²⁰.

¹⁸ Anderson, "Debating the Long Tail."

¹⁹ Ibid.

²⁰ Elberse and Oberholzer-Gee, "Superstars and Underdogs."

Integration and Specification

As early as 2008, researchers begin attempting to bridge the gap between the competing viewpoints of long tail and superstar emergence. They do so by building models, both conceptual and, later, empirical, supporting an integrated view of these distributions in online markets. One of the first conceptual models proposes a rigorous approach to identifying some “determinants of superstar and long tail effects.” Using neoclassical microeconomic profit models, researchers show that a reduction in “variable costs relative to fixed costs support superstar[s]” and that a reduction in “fixed costs relative to variable costs support long tail[s]”²¹. Little research that follows this work relies on a similar model, but it illustrates an early impetus for a model that can predict the emergence of long tails or superstars, without claiming one is more important than the other.

A 2011 evaluation of a monopolistic information-good market advances a compelling model of both supply- and demand-side technological and non-technological factors heretofore absent from long tail and superstar research. In it, Hinz, Eckert, and Skiera observe what many other researchers seem to ignore—the products in the middle of the distribution. They discuss five elements—search technology, assortment size²², consumer demographics, quality of overall assortment, and seasonal effects—that they show can help predict the emergence of long tails or superstars. Their research suggests that assortment size and the quality of the overall assortment are helpful parameters in predicting outcomes in all of the categories they study: demand per customer, share of purchased products,

²¹ Kendall and Tsui, “The Economics of the Long Tail.”

²² Also, “product variety.”

and demand distribution across products. They also posit that the effect of search technology on the market distribution is dependent on the assortment size. A greater variety of products in the market with search filters or recommendation network(s) corresponds to an uptick in blockbuster sales; less variety corresponds to greater sales of niche products²³. This finding, echoed in a later evaluation of the software download market, supports the prior assertion that Anderson's condition on the presence of search technology needs strong conceptual specification to hold²⁴.

Synthesis

The literature on the emergence of long tails up to this point has developed a predictive macro model of market distributions. Much of the work drew upon research in particular markets in order to infer conclusions about the emergence of long tails and superstar effects in all markets. These macro-level models will be useful as this paper progresses to evaluate the book publishing industry. In particular, the aforementioned assumptions of product variety, consumer informedness, and assortative mixing will be explored further in Chapter Four. A survey of the contemporary book publishing industry will discuss two additional micro-level assumptions for the industry-specific research design I expand upon in the fourth chapter, as well.

²³ Hinz, Eckert, and Skiera, "Drivers of the Long Tail Phenomenon."

²⁴ Zhou and Duan, "Online User Reviews, Product Variety, and the Long Tail."

CHAPTER THREE

Survey of the U.S. Book Publishing Industry

The United States book publishing industry is competitive, diverse, and always changing. “IBISWorld Industry Reports” defines the industry as one in which, “[publishers] edit and design books [and] engage in marketing activities [as well as] make distribution agreements with wholesalers.” Importantly, the report notes that “authors or print[ers] are not included as part of the market²⁵.” This limitation excludes strictly print-on-demand (POD) companies and authors publishing books for only personal use. However, this does not exclude companies providing self-publishing resources, which have become more and more relevant with the technological change occurring in the market.

In the advent of the digital age and an economy in flux, ebooks and the amount of college students prove to be two large forces in the redistribution and restructuring of the publishing market. As consumers move to purchase content in coming years, it will be important for book publishers to pay attention to the industry’s nuances. A brief history of twentieth and early twenty-first century developments, an evaluation of contemporary industry structure, and an exploration of the two forces mentioned above should clarify those nuances for further contemplation within this paper’s market distribution model.

²⁵ Kaczanowska, *Book Publishing in the US*.

History

In the early 1900s, “the [book publishing] industry was insular, self-conscious of its traditions, and [...] unlikely to experiment with new business practices²⁶.” It followed rigid, mechanical processes that took content in the form of manuscripts through editing, presses, publisher warehouses, and then to the market. Not until after 1945, following World War II and the return of American soldiers, did the industry need to begin adapting to a rapidly shifting state of demand.

Soldiers and a newly interested consumer base injected the market with never-before-seen demand in locations far out of the reach of traditional booksellers. Though publishers held tight to a certain level of quality in the content they released through their imprints, the industry restructured to accommodate for these new customers with wider physical distribution and greater product variety. This “paperback revolution” sought to reach more consumers with a taste for the nontraditional. Numerous new imprints and publishers formed to grab a piece of the profit pie. In the meantime, vanity presses that allowed authors to self-publish and print on demand, such as Vanguard, continued struggling to find demand.

In the 1960s, corporate media conglomerates began to invest in the publishing industry by acquiring and combining as many publishers as possible. This phenomenon was short-lived, however. Due to unpredictable sales cycles, book industry dynamics proved too unruly for the business executives concerned solely with the bottom line; by the 1970s, many of the collected imprints were sold

²⁶ Wilson and Lucyk, *A Very Brief History of the Book Publishing Industry*.

off or left to die. Nevertheless, international conglomerates began to form in lieu of the corporate publisher by the late 1970s. Many of these still exist today, in the form of companies like Penguin, Random House, and HarperCollins.

Throughout the century, the retail structure of book sales changed enormously, as well. Following World War II, the sale of books became less concentrated in urban, metropolitan areas and began to spread into general stores and other widespread retail entities. Soon after, suburban retail chains began to emerge and aggressively expand—names like Barnes & Noble and Borders became synonymous with consumers' ideas of good, quality books. This retail transformation allowed a greater variety of consumers to gain access to the published content market, creating wider diversity of demand.

More recently, the market has expanded even further. The technological advances discussed by Anderson and others catalyzed the formation of new self-publishing, print-on-demand, and niche content providers unlike any other shift had previously. The “Digital Revolution” contributed greatly to the structure of the book publishing industry of today.

Contemporary Structure

Both traditional and new (internet-enabled) publishing industries find initial value in content created by writers. Traditional markets require that writers' content filter through literary agents and publishers before reaching any consumers. In general, these two points act as gatekeepers for content of questionable marketable value. Agents and publishers receive large amounts of content from writers and, often, cannot filter all of it; much of this ends up in what

these entities call “slush piles,” which often go unread for months and rarely end up in the market. This is a sad reality for most content that attempts to navigate through traditional publishing channels, which focus on products known or assumed to be bestselling material. Writers need not worry, though, for around 83 percent of publishers do not fall into this highly selective, discriminating category²⁷.

Apart from the lucrative niche of textbook publishing, which is very concentrated (with only about four significant players), the concentration level of the book publishing industry as a whole is low. In addition, competition is high and predicted to remain so²⁸. As new formats for content emerge, like ebooks, increasing numbers of books are published and sold through online channels; currently, about twenty-five percent of the entire published content market is online. This figure includes sales through companies like Amazon and dynamic print on demand (POD) services.

Also part of this diverse market is the self-publishing phenomenon of “author services” business models. Some say the self-publishing sector is now “a viable route for [niche authors]” but that those titles may still have difficulty finding places amongst content published through traditional channels. Notwithstanding, self-publishing is a rising trend. In a recent report, Bowker, the company that creates and assigns ISBNs in the U.S., notes a 287 percent increase in the number of self-published titles since 2006. The company says that “148,424 print books were self-published in 2011,” which adds up to “43 percent of all print books published in the U.S.” in the same year. In addition, “87,201 ebooks were self-published” with

²⁷ Kaczanowska, *Book Publishing in the US*.

²⁸ Ibid.

registered ISBNs that year. The publishing industry is transforming rapidly, and companies must begin to innovate in order to compete with companies “once relegated to the outskirts of [the market].”

Ebooks and College Students

The digital book format, or ebook, is becoming more and more popular for self-publishers and traditional publishers, alike. From 2010 to 2011, Bowker records a 129 percent increase in the amount ebook ISBNs created through self-publishing. Other sources predict that 2013 will see ebooks command a 45 percent share of the “overall trade book sales”; the publishing market is made up of nearly a majority of digital content. Even for companies selling niche content alongside blockbusters, like Amazon, ebooks are selling “105 [...] for every 100 printed” books²⁹. In fact, from December 2011 to January 2012 the percentage of Americans with at least one digital reading device jumped by eleven, from eighteen to twenty-nine, illustrating the possibility of increasing ebook demand with which U.S. publishers must contend³⁰. Some estimates go so far as to say ebooks will claim 90 percent of the market by 2030³¹. Whatever the future holds, research suggests the current market distribution is affected by the emergence of ebooks.

A steady increase in the number of college students is also a factor in the determination of book publishing market structure. Enrollment increased 11 percent from 1990 to 37 percent from 2000 to 2010³². The Fall Semester of 2011

²⁹ Stone, “Amazon’s Hit Man.”

³⁰ Ibid.

³¹ Hua, Cheng, and Wang, “Electronic Books.”

³² *Postsecondary Graduation Rates*.

saw a 0.7 percent increase in enrollment over the previous year, but the amount of students enrolling in 2012 decreased by 1.8 percent³³. Despite this modest decline, the demand for textbooks created by this large college population fuels the book publishing industry quite comfortably. This sector is the largest contributor to industry revenues, independently “generating about 28.5 percent of revenue³⁴.”

These students are, in general, finishing school, as well. The most recent data shows that 58 percent of first-time enrollees starting school in 2004 completed an undergraduate degree by 2010³⁵. This was an increase from a similar group, observed from 1996 to 2002, which graduated 55 percent of the first-time enrollees. More college graduates leads to greater demand for published materials, overall, because those who complete a university education tend to read consistently for the rest of their lives³⁶.

Synthesis

The U.S. book publishing industry’s dynamic market structure will likely face multiple changes growing out of technology and new entrants in the coming five years. The online portion of the market is, no doubt, composed of a great amount of niche content, but it is also home to flattening economies of scale³⁷. Even printed blockbusters become more popular online not only for consumers’ innate interest in talent or quality, as Rosen would suggest, but for cheaper models of production of that quality content. Increasing production of blockbuster content by traditional

³³ Lederman, “College Enrollments Fell This Fall.”

³⁴ Kaczanowska, *Book Publishing in the US*.

³⁵ *Postsecondary Graduation Rates*.

³⁶ Kaczanowska, *Book Publishing in the US*.

³⁷ International Conference on the Book (3rd : 2005 : Oxford, Cope, and Phillips, *The Future of the Book in the Digital Age*.

publishers may lead to heavier concentration in the head of the market distribution. Compounding this, greater textbook sales, which benefit the players at the head of the market, as well, might contribute to this concentration. On the other hand, greater production of ebooks coupled with heightened demand for reading materials from college graduates in the market may push the distribution in the opposite direction—toward a longer, wider tail. In either case, researchers and managers have compelling reasons to invest resources into better understanding the effects of these changes.

CHAPTER FOUR

Proposed Content Model

Introduction to the Directed Acyclic Graphical Model

The conditions heretofore introduced relating to the emergence of long tails and superstars in online markets are here considered as part of a proposed model for further research. To construct this model, I utilize the directed acyclic graphic (DAG) framework created by Judea Pearl and adapted by Stephen Morgan and Christopher Winship in *Counterfactuals and Causal Inference* (2007). Below, Figure 4.0 illustrates Anderson’s conditions for long tail emergence in a DAG. We shall use this as a simple example in order to explain the mechanics of a DAG.

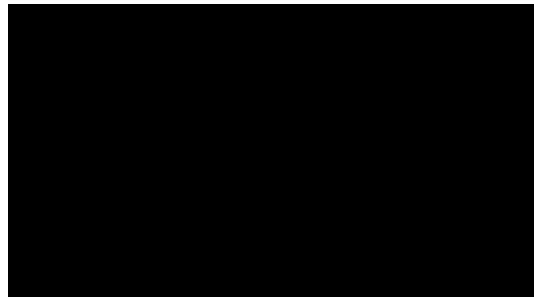


Figure 4.0

In any DAG, “each node represents an observable random variable,” and that “the variable at the origin of a directed edge causes the variable at the terminus of the [edge].” So, this DAG implies that new technology has a causal effect on the disintermediation of production, the disintermediation of distribution, and the filtration mechanisms present in the market. This would also suggest that, using

proper statistical procedures, a qualified causal effect could be derived for each relationship of interest present in the DAG: $P \rightarrow Y$, $D \rightarrow Y$, and $S \rightarrow Y$ are our relationships of interest.

The DAG assumes that the relationships depicted are, in fact, “theoretical propositions that have achieved consensus in the relevant scholarly community.” This assumption does not hold for this particular DAG, as the literature review has shown, so we will continue to expand upon it throughout this chapter. In actuality, the assumption never really holds because, as Morgan and Winship note, “[o]ther causes [of the relationships] are left implicit” in almost all cases; this is why there is room left for statistical error and why each node is not considered a definite observation, but instead a random variable with “a probability distribution³⁸.”

An important feature of the DAG framework is its “acyclic” designation. This term simply implies that a proper DAG cannot contain any cycles of causation. This rule will become more obvious and important as we proceed through the analysis of the relationships in this model.

In order to extract the aforementioned relationship of interest $P \rightarrow Y$, for example, we shall use Pearl’s strategy for conditioning on variables. This is an exercise in identifying other paths of causality, which Pearl calls “backdoor paths,” that serve to confound the relationship of interest. There are two such backdoor paths in this DAG, namely $P \leftarrow T \rightarrow D \rightarrow Y$ and $P \leftarrow T \rightarrow S \rightarrow Y$. Using statistical techniques or simpler matching methods, one can “block” these backdoor paths and

³⁸ Morgan and Winship, *Counterfactuals and Causal Inference*.

determine a more accurate, less biased probable $P \rightarrow Y$ relationship. In order to do so effectively, Pearl's "Backdoor Criterion" must be met.

Morgan and Winship explain that "the causal effect [for the relationship of interest] is identified if and only if all backdoor[s] are conditioned " where there is a "chain of mediation," like $T \rightarrow D \rightarrow Y$, or a "fork of mutual dependence," like $P \leftarrow T \rightarrow D$, but *never* where there is a "collider," like D in $E \rightarrow D \leftarrow T$. Conditioning on T in our example will block both backdoor paths and, if all assumptions were in place, allow us to extract the causal relationship of interest, $P \rightarrow Y$. Trying this technique with $D \rightarrow Y$ and $S \rightarrow Y$ will yield the same result. A square around the conditioned variable is acceptable notation.

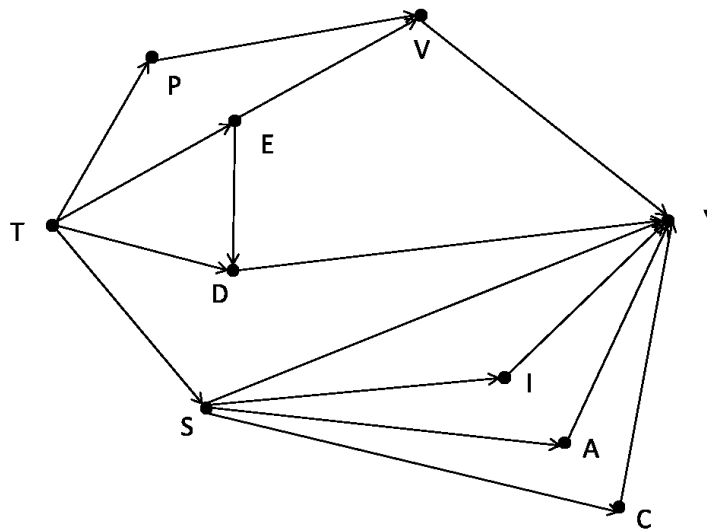


Figure 4.1

The DAG shown above contains the variables listed below in Table 4.0. This is the most expansive the model will be for our purposes, although there would be room for additional variables in a fuller evaluation of the market. I focus on five new variables for the sake of brevity.

| Variable | Description |
|----------|--|
| Y | Likelihood of Long Tail or Superstar emergence in the Book Publishing Industry |
| T | New technology (Internet-enabling) |
| P | Disintermediation of Production within the Book Publishing Industry |
| V | Product Variety |
| D | Disintermediation of Distribution within the Book Publishing Industry |
| E | Ebook presence in the consumer market for content |
| S | Supply Filter(s) within the market |
| I | Higher degree of Consumer Informedness |
| A | Assortative Mixing |
| C | Number of College Students |

Table 4.0

As shown above, there are many more backdoor paths to the three basic relationships of interest than we began with on Anderson’s conditions, alone. We shall explore each of the added variables individually in the following sections both conceptually and within the DAG framework.

Product Variety (V)

As discussed in Chapter Two, product variety is an implicit condition in Anderson’s original model that research suggests should be considered independently. In this rendering, the model gains the additional complication of a new chain of mediation, $P \rightarrow V \rightarrow Y$. Conditioning on T will block the existing two backdoor paths mentioned above, but should we condition on V, as well, to extract the $P \rightarrow Y$ relationship?

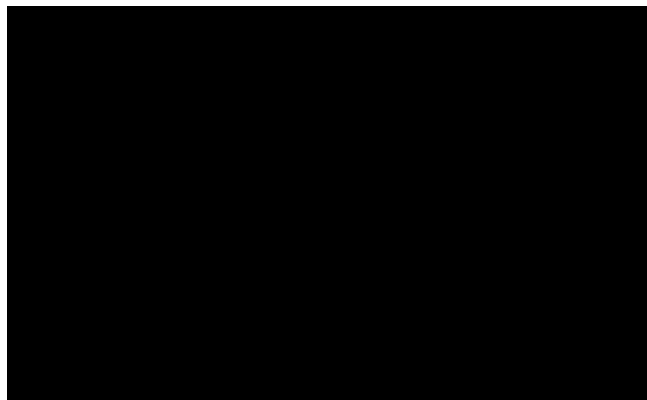


Figure 4.2

In conditioning on V , we remove the effect that product variety contributes to the emergence of either distribution. The literature suggests that this conditioning removes a significant portion of the effect that P could possibly have on Y . In fact, it argues that increased product variety contributes directly to the emergence of a long tail. However, conditioning on this random variable will disallow the $P \rightarrow Y$ relationship extraction in the above DAG, according to Morgan and Winship.

The placement of V in this DAG conforms to the literature in that multiple articles empirically show at least a correlation between product variety and long tail emergence ($P \rightarrow Y$). In addition, Anderson and others show that, absent constraints on the democratization of production, product variety is quite inevitable ($P \rightarrow V$).

Further empirical research could attempt to evaluate a situation wherein a market features democratized production but where product variety is artificially limited. If the literature's suggestions are in line with reality, then results should show that disintermediation of production is a weak condition for long tail emergence unless it leads to product variety.

Ebook Emergence (E)

As discussed in Chapter Three, the emergence and ownership of ebooks in the consumer market contributes in various ways to changing the US book publishing industry. We focus on the E as it affects $D \rightarrow Y$ in this example. Although we could use the same conditioning technique as above (blocking on T and V), we block on T and E . This should highlight the effect disintermediation of distribution

has upon the emergence of a long tail/superstar within the US publishing market, independent of the effect ebooks have.

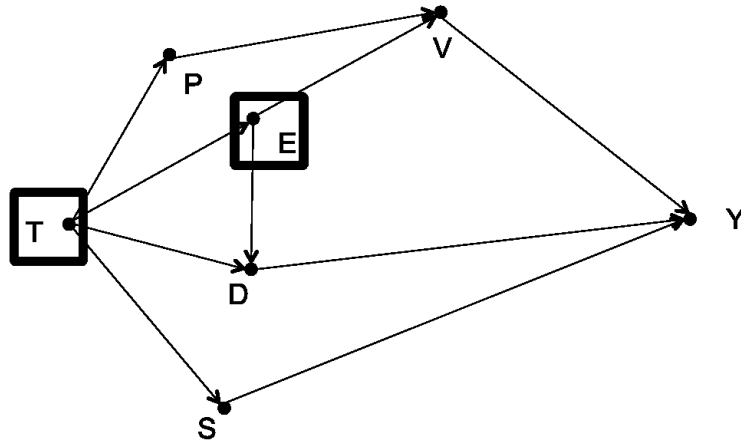


Figure 4.3

Conditioning on E would allow researchers to focus on whether D tends to lead toward a long tail or a superstar effect, assuming omitted variables do not create bias in the estimation. The placement of E, as well as its relationships within the DAG, conforms to the literature’s claims. Ebook formats significantly decrease the costs of individual book publishing ($E \rightarrow D$) and allow more producers to create content that may have otherwise gone unpublished for prohibitive costs ($E \rightarrow V$). Internet-enabling technology is pivotal to the functionality of ebooks, as well ($T \rightarrow E$).

Further empirical research could evaluate a market such as Amazon prior to the release of the Kindle for evidence that $D \rightarrow Y$ could hold without E. Anderson may touch on this, but it is unclear whether the Amazon data in his book includes both ebooks and traditional books or only one at the exclusion of the other.

Assortative Mixing Within Categories (A)

As discussed in Chapter Two, assortative mixing within categories, or the deliberate presentation of products by a passive supply filter to the consumer,

promotes the emergence of a long tail. In this rendering of the DAG, the relationship of interest is $S \rightarrow Y$. We will continue to block on T and will now block on A, too.

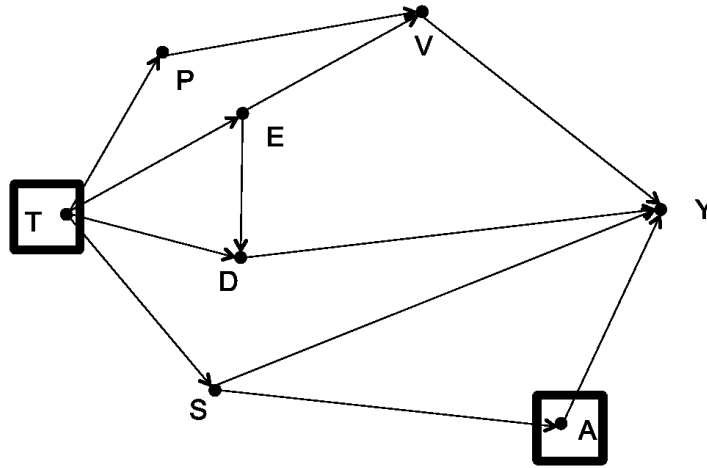


Figure 4.4

Blocking the effect of $A \rightarrow Y$ is part of a sequence of conditioning that should highlight the $S \rightarrow Y$ relationship. This particular conditioning would simply remove a particularly popular method of passive supply filtration from consideration.

Amazon’s recommendation systems are a great example of this filtration mechanism. Often, when looking to purchase, for example, a test preparation book, Amazon will suggest another book designed to prepare for the same test or one meant to assist test-takers in various other ways.

The placement of A in the DAG is present in the literature, which claims A leads to a market that is “more evenly distributed” ($A \rightarrow Y$)³⁹. Also, A is a type of S that has a supposed independent effect on Y ($S \rightarrow A$). Conditioning on A should, if the literature’s arguments hold, diminish the long tail. An empirical research could simply consider a market not using such passive supply filters.

³⁹ Oestreicher-Singer and Sundararajan, “Recommendation Networks and the Long Tail of Electronic Commerce.”

Higher Degree of Consumer Informedness (I)

Also discussed in Chapter Two and also in relation to the relationship $S \rightarrow Y$, the importance of consumer informedness as a product of supply filters is illustrated empirically in the literature. We progress from the previous example with the same blocked variables, but we now block I, as well.

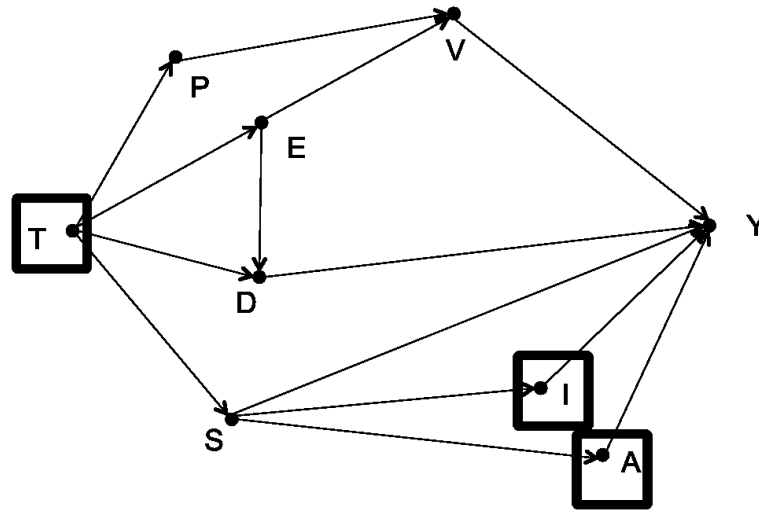


Figure 4.5

Blocking on I is the second mechanism in the sequence of conditioning meant to extract the $S \rightarrow Y$ relationship. If the literature's findings are correct, then the removal of I from this relationship should contribute to a less prominent long tail; it may have plenty of products, but they will not sell as well as they would with more informed consumers. In essence, blocking on I would require a market with supply filters dedicated to leading consumers towards particular products rather than filters designed to promote discovery.

The placement of I in the DAG follows the same logic as A; the literature shows it as having a direct effect on promoting long tail emergence ($I \rightarrow Y$), and I is an element of S that has a supposed independent effect on Y ($S \rightarrow I$). Further

empirical research in this area should, in addition to what is prescribed above for A, regard a market wherein consumers are not informed at all by the supply filtration mechanisms therein. This could be achieved only in a market lacking passive filters, which are by their nature, informative.

Number of College Students (C)

Discussed in Chapter Three, the number of college students in the US is the final consideration within this model and the last mechanism in the sequence of conditioning for extracting the $S \rightarrow Y$ relationship. Not only do we block T, A, and I as in the prior example, we must now condition upon C.

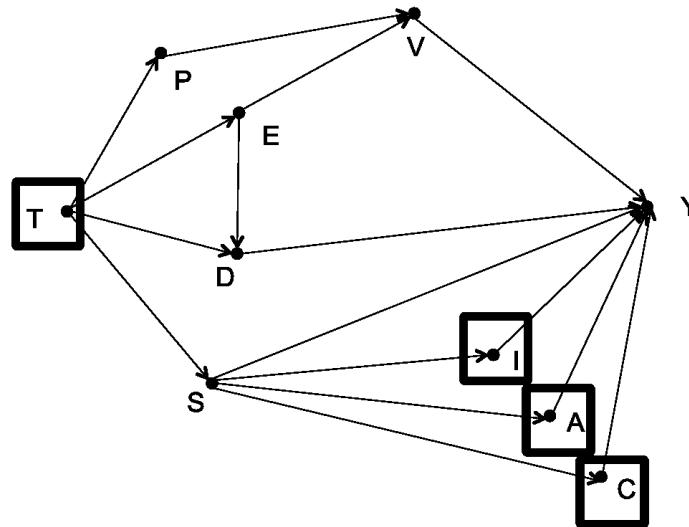


Figure 4.6

The data presented in the literature shows that increasing enrollment at colleges and universities in the US increases textbook sales; the largest four “superstar” companies in the book publishing industry publish almost all textbooks. The implication is that increasing amounts of college students will lead to a superstar effect in the US book publishing industry ($C \rightarrow Y$). Profitable textbook sales

(new or used) incentivize producers to create supply filters geared toward college interests ($S \rightarrow Y$).

Further empirical research can observe this step of the conditioning sequence by evaluating sectors of the industry that produce no textbooks.

CHAPTER FIVE

Application

Assume two managers in an existing firm in the US book publishing industry are attempting to make a decision on whether to follow Anderson's directive to "sell less of more." They manage a firm that publishes content within the US in print and digitally; currently, content published through the firm is subject to quality control and extensive editing processes. The firm's website allows consumers to search for content with active search filters and provides no passive filtration system.

The managers wish to know if (i) eliminating distributors from the firm's to-market supply chain, (ii) providing self-publishing resources to writers, and (iii) improving the firm's website search filters would be effective in facilitating more fluid, diverse product sales for their firm. The three relationships of interest for the managers are represented in the model (articulated in Chapter Four) as follows: (i) $D \rightarrow Y$; (ii) $P \rightarrow Y$; and (iii) $S \rightarrow Y$.

Selecting and collecting the proper data is essential to the operation of the model. Perhaps most essential variable to observe is T. Without considering "technology" an endogenous observable random variable, testing the relationships within the model becomes tenuous. Considering the nature of the book publishing

industry and the nature of Anderson's assumptions, a source of data for T could be the amount of firm server space. This source is observable over time and would already be in the company's records. Server space is determinative of network limitations and, thus, the company's ability to store and manage content.

Blocking on T using the server space data is sufficient to test the managers' second consideration, whether providing self-publishing resources to writers would result in long tail or superstar effects. This helps the managers to avoid heavy investment in self-publishing resources that may be unpopular or, in contrast, encourages entrance into the self-publishing market where the model shows a possibly vibrant long tail.

In order to test their first idea, the elimination of current intermediaries in the distribution chain, the managers will need to block on T and E. The server space data may be used here, as well, to observe T. Data for E may be gleaned from the firm's past and current sales of ebooks. For a firm without a digital publishing division, observation of V would be necessary to properly block all backdoor paths. Controlling for variation in these data sources should provide the managers with an estimation of the effect of eliminating their intermediary partners without having to blindly sever existing business relationships.

For the final consideration, improving the website search filters, the managers must provide data for to block on T, I, A, and C. Data on C is readily available online from federal documents detailing number of college enrollees. The firm already blocks on A by providing no passive search filters. Data on consumer purchases observed following the removal of intra-genre content recommendation

systems would be sufficient for blocking on A for a company using such passive filtration. It is perhaps most difficult to collect data on, as it is a measurement of the change in informedness of consumers purchasing content. A voluntary survey given to consumers upon completion of an order asking about their knowledge of the product on a numerical scale could be a workable approach. The quantitative data collected from such a survey could allow the managers to utilize the model appropriately.

After blocking on the proper variables and observing the relationships of interest, the literature would suggest that the managers should see some semblance of both the long tail and superstar effects within their market.

A complete model of long tail and superstar emergence is beyond the scope of this paper, but the benefits of such a model for businesses operating online and for academic research are many and valuable. This paper hopes to have further integrated the previous literature's considerations in a convenient model for future exploration into the emergence of long tails and superstars in microeconomic industries.

BIBLIOGRAPHY

- Anderson, Chris. "Debating the Long Tail." *Harvard Business Review*. Accessed March 17, 2013. http://blogs.hbr.org/cs/2008/06/debating_the_long_tail.html.
- . *The Long Tail: Why the Future of Business Is Selling Less of More*. 1st ed. New York: Hyperion, 2006.
- . "Wired 12.10: The Long Tail." Accessed March 8, 2013. <http://www.wired.com/wired/archive/12.10/tail.html>.
- Black Book - IP Video: The Push to 80%*. BernsteinResearch, n.d.
- Brynjolfsson, Erik, Yu Jeffrey Hu, and Michael D. Smith. "From Niches to Riches: Anatomy of the Long Tail." *SSRN eLibrary*. Accessed December 12, 2012. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=918142.
- Brynjolfsson, Erik, Yu Hu, and Michael D. Smith. "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers." *Management Science* 49, no. 11 (November 2003): 1580–1596.
- . "Long Tails Vs. Superstars: The Effect of Information Technology on Product Variety and Sales Concentration Patterns." *Information Systems Research* 21, no. 4 (December 2010): 736–747.
- Clemons, Eric K., and Guodong (Gordon) Gao. "Consumer Informedness and Diverse Consumer Purchasing Behaviors: Traditional Mass-market, Trading down, and Trading Out into the Long Tail." *Electronic Commerce Research and Applications* 7, no. 1 (Spring 2008): 3–17. doi:10.1016/j.elerap.2007.10.001.
- Crowdhry, Trip. *Internet As A Building Block*. Transcript, January 25, 2010.
- Elberse, A, and F Oberholzer-Gee. "Superstars and Underdogs: An Examination of the Long Tail Phenomenon in Video Sales." *MSI Reports* (2007). http://www.people.hbs.edu/aelberse/papers/hbs_07-015.pdf.
- Fleder, Daniel, and Kartik Hosanagar. *Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity*. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, April 17, 2007. <http://papers.ssrn.com/abstract=955984>.

- Hinz, Oliver, Jochen Eckert, and Bernd Skiera. "Drivers of the Long Tail Phenomenon: An Empirical Analysis." *Journal of Management Information Systems* 27, no. 4 (Spring 2011): 43–69.
- Hua, Guowei, T.C.E. Cheng, and Shouyang Wang. "Electronic Books: To 'E' or Not to 'E'? A Strategic Analysis of Distribution Channel Choices of Publishers." *International Journal of Production Economics* 129, no. 2 (February 2011): 338–346. doi:<http://dx.doi.org/10.1016/j.ijpe.2010.11.011>.
- International Conference on the Book (3rd : 2005 : Oxford, England), Bill Cope, and Angus Phillips. *The Future of the Book in the Digital Age*. Oxford: Chandos,, 2006.
- Kaczanowska, Agata. *Book Publishing in the US*. IBISWorld Industry Report. Book Publishing the US. IBISWorld, February 2013. <http://clients1.ibisworld.com/reports/us/industry/default.aspx?entid=1233>
- Kendall, Todd D., and Kevin Tsui. "The Economics of the Long Tail." *B.E. Journal of Economic Analysis & Policy: Topics in Economic Analysis & Policy* 11, no. 1 (January 2011): 1–18.
- Lederman, Doug. "Fall 2012 College Enrollments Fell by 1.8% | Inside Higher Ed." *Inside Higher Ed*, December 19, 2012. <http://www.insidehighered.com/news/2012/12/19/fall-2012-college-enrollments-fell-18>.
- Lovelock, John-David. *Forecast Alert: IT Spending, Worldwide, 4Q12 Update | 2291618*. Market Analysis and Statistics. Forecast Alert. Gartner, January 2, 2013. <http://www.gartner.com/DisplayDocument?id=2291618>.
- McPhee, William N. *Formal Theories of Mass Behavior*. [New York]: Free Press of Glencoe, 1963.
- Morgan, Stephen L., and Christopher Winship. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, 2007.
- Oestreicher-Singer, Gal, and Arun Sundararajan. "Recommendation Networks and the Long Tail of Electronic Commerce." *MIS Quarterly* 36, no. 1 (March 2012): 65–A4.
- Peltier, Stephanie, and Francois Moreau. "Internet and the 'Long Tail Versus Superstar Effect' Debate: Evidence from the French Book Market." *Applied Economics Letters* 19, no. 7–9 (June 2012): 711–715.
- Rosen, Sherwin. "The Economics of Superstars." *American Economic Review* 71, no. 5 (December 1981): 845.

- “Shirky: Power Laws, Weblogs, and Inequality.” Accessed July 20, 2012.
http://www.shirky.com/writings/powerlaw_weblog.html.
- Stone, Brad. “Amazon’s Hit Man.” *BusinessWeek: Magazine*, January 25, 2012.
<http://www.businessweek.com/magazine/amazons-hit-man-01252012.html>.
- The Condition of Education - Postsecondary Education - Completions - Postsecondary Graduation Rates - Indicator 45 (2012)*, 2012.
http://nces.ed.gov/programs/coe/indicator_pgr.asp.
- Wilson, Britanie, and Jeremy Lucyk. *A Very Brief History of the Book Publishing Industry*. Kindle. Entering the Publishing Industry. Centennial College Press, n.d.
- Zhou, Wenqi, and Wenjing Duan. “Online User Reviews, Product Variety, and the Long Tail: An Empirical Investigation on Online Software Downloads.” *Electronic Commerce Research & Applications* 11, no. 3 (June 2012): 275–289.